```
FT   MOD_RES     1057   1057       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CONFLICT      25     25       P -> T (IN REF. 1).
FT   CONFLICT     679    679       G -> D (IN REF. 3).
FT   CONFLICT     783    784       LV -> VL (IN REF. 1).
FT   CONFLICT     917    917       S -> C (IN REF. 1).
FT   CONFLICT    1341   1367       QLTSCLNGSGPVPAPPPTPGNHERGAA -> RSPPV
FT                                 (IN REF. 3).
SQ   SEQUENCE   1367 AA;  152516 MW;  EFC99704F1DCA266 CRC64;

  Query Match             4.7%;  Score 109;  DB 1;  Length 1367;
  Best Local Similarity   21.0%;  Pred. No. 12;
  Matches   45;  Conservative   36;  Mismatches   73;  Indels   60;  Gaps    9;

Qy          9 HLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSF 68
              |::   :: : |||    :       ::| | ||| :       |        ||:| :
Db        544 HVIRGPEITVQPAAQPTEQ-------ESVSLLCTADRNTFENLTWYK------LGSQATS 590

Qy         69 LTKGPSKLNDRADSRRSLW--------DQGNFPLII--KNLKIEDSDTYICEVEDQKEE- 117
              :  | | |     :  :||       :  |  ||:  :|  ::|   |:|  :|:| :
Db        591 VHMGES-LTPVCKNLDALWKLNGTMFSNSTNDILIVAFQNASLQDQGDYVCSAQDKKTKK 649

Qy        118 -----VQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKN---------- 162
                  ||::    |   | |: |: |:      | :  | |  |   |
Db        650 RHCLVKQLIILERMAPMITGNLENQTTTI------GETIEVTC--PASGNPTPHITWFKD 701

Qy        163 ------------IQGGKTLSVSQLELQDSGTWTC 184
                         | : |:: ::  :| | :||
Db        702 NETLVEDSGIVLRDGNRNLTIRRVRKEDGGLYTC 735


RESULT 72
PVR2_HUMAN
ID   PVR2_HUMAN     STANDARD;      PRT;   538 AA.
AC   Q92692; O75455; Q96J29;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Poliovirus receptor related protein 2 precursor (Herpes virus entry
DE   mediator B) (HveB) (Nectin 2) (CD112 antigen).
GN   PVRL2 OR PRR2 OR HVEB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM DELTA).
RX   MEDLINE=95347610; PubMed=7622062;
RA   Eberle F., Dubreuil P., Mattei M.-G., Devilard E., Lopez M.;
RT   "The human PRR2 gene, related to the human poliovirus receptor gene
RT   (PVR), is the true homolog of the murine MPH gene.";
RL   Gene 159:267-272(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA).
RX   MEDLINE=98321161; PubMed=9657005;
RA   Warner M.S., Geraghty R.J., Martinez W.M., Montgomery R.I.,
RA   Whitbeck J.C., Xu R., Eisenberg R.J., Cohen G.H., Spear P.G.;
RT   "A cell surface protein with herpesvirus entry activity (HveB) confers
RT   susceptibility to infection by mutants of herpes simplex virus type
RT   1, herpes simplex virus type 2, and pseudorabies virus.";
RL   Virology 246:179-189(1998).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 31-538 FROM N.A.
RA   Yoshiura K., Murray J.C.;
RT   "A transcriptional map in the region of 19q13 derived using direct
RT   sequencing and exon trapping.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 449-538 FROM N.A.
RX   MEDLINE=99449047; PubMed=10520737;
RA   Freitas E.M., Zhang W.J., Lalonde J.P., Tay G.K., Gaudieri S.,
RA   Ashworth L.K., Van Bockxmeer F.M., Dawkins R.L.;
RT   "Sequencing of 42kb of the APO E-C2 gene cluster reveals a new gene:
RT   PEREC1.";
RL   DNA Seq. 9:89-101(1998).
CC   -!- FUNCTION: RECEPTOR FOR ALPHAHERPESVIRUS (HSV-1, HSV-2 AND
CC       PSEUDORABIES VIRUS) ENTRY INTO CELLS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Delta;
CC         IsoId=Q92692-1; Sequence=Displayed;
CC       Name=Alpha;
CC         IsoId=Q92692-2; Sequence=VSP_002628, VSP_002629;
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC   -!- DATABASE: NAME=PROW; NOTE=PROW 1:74-77(2000);
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/guide/204270028_g.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X80038; CAA56342.1; -.
DR   EMBL; AF058448; AAC23797.1; -.
DR   EMBL; BC003091; AAH03091.1; -.
DR   EMBL; AF044968; AAC82348.1; -.
DR   EMBL; AF044962; AAC82348.1; JOINED.
DR   EMBL; AF044963; AAC82348.1; JOINED.
DR   EMBL; AF044964; AAC82348.1; JOINED.
DR   EMBL; AF044966; AAC82348.1; JOINED.
DR   EMBL; AF044967; AAC82348.1; JOINED.
DR   EMBL; AF050154; AAD02503.1; -.
DR   PIR; I68093; I68093.
DR   Genew; HGNC:9707; PVRL2.
DR   MIM; 600798; -.
DR   GO; GO:0005886; C:plasma membrane; TAS.
DR   GO; GO:0015026; F:coreceptor activity; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00409; IG; 1.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain; Receptor; Transmembrane; Glycoprotein; Signal;
KW   Repeat; Alternative splicing.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    538       POLIOVIRUS RECEPTOR RELATED PROTEIN 2.
FT   DOMAIN       32    360       EXTRACELLULAR (POTENTIAL).
```

```
FT   TRANSMEM     361     381       POTENTIAL.
FT   DOMAIN       382     538       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        32     156       IG-LIKE V-TYPE.
FT   DOMAIN       162     256       IG-LIKE C2-TYPE 1.
FT   DOMAIN       261     345       IG-LIKE C2-TYPE 2.
FT   DISULFID      54     140       BY SIMILARITY.
FT   DISULFID     183     238       BY SIMILARITY.
FT   DISULFID     283     329       BY SIMILARITY.
FT   CARBOHYD     137     137       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     324     324       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC     351     479       NTAGAGATGGIIGGIIAAIIATAVAATGILICRQQRKEQTL
FT                                  QGAEEDEDLEGPPSYKPPTPKAKLEAQEMPSQLFTLGASEH
FT                                  SPLKTPYFDAGASCTEQEMPRYHELPTLEERSGPLHPGATS
FT                                  LGSPIP -> RASPRDVGPLVWGAVGGTLLVLLLLAGGSLA
FT                                  FILLRVRRRRKSPGGAGGGASGDGGFYDPKAQVLGNGDPVF
FT                                  WTPVVPGPMEPDGKDEEEEEEEEKAEKGLMLPPPPALEDDM
FT                                  ESQLDGSLISRRAVYV (in isoform Alpha).
FT                                  /FTId=VSP_002628.
FT   VARSPLIC     480     538       Missing (in isoform Alpha).
FT                                  /FTId=VSP_002629.
SQ   SEQUENCE   538 AA;  57742 MW;  3AE4F83E92F6F624 CRC64;

  Query Match             4.6%;  Score 108;  DB 1;  Length 538;
  Best Local Similarity   21.3%;  Pred. No. 4.4;
  Matches  112;  Conservative   58;  Mismatches  204;  Indels  152;  Gaps   25;

Qy          2 NRGVPFRHLLLVLQLALLPAATQGNKV-----VLGKKGDTVELTC--------------- 41
              :|  |   | :| | ||    | :|     | |: | |||| |
Db         10 SRSPPTPLLWPLLLLLLLETGAQDVRVQVLPEVRGQLGGTVELPCHLLPPVPGLYISLVT 69

Qy         42 ------TASQKKSIQFHWKNSNQI---KILGNQGSFLTKGPSKLNDRADSRRSLWDQGNF 92
                  |: :    || |         |    : ||::   :| :   |:   | |
Db         70 WQRPDAPANHQNVAAFHPKMGPSFPSPKPGSERLSFVS---AKQSTGQDTEAELQDA--- 123

Qy         93 PLIIKNLKIEDSDTYICEVED-QKEEVQ----LLVFGLTANSDTHLLQGQSLTLTLESPP 147
               | :  | :||   | ||      |  |:    | |      |       | ::  :|    |
Db        124 TLALHGLTVEDEGNYTCEFATFPKGSVRGMTWLRVIAKPKN------QAEAQKVTFSQDP 177

Qy        148 GSSPSVQCRSPRGK---NIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAF 204
               :    | |  |:    |    :|     | | |||   ||      |  :  :|
Db        178 --TTVALCISKEGRPPARISWLSSLDWEAKETQVSGTLAGTV-----TVTSRFTLVPSGR 230

Qy        205 QKASSIVYKKEGEQVE--FSFPLAFTVE-----KLTGSGELWWQAE---------RASSS 248
              ::  | | |  |      |: :|       ::|  : |:             |::
Db        231 ADGVTVTCKVEHESFEEPALIPVTLSVRYPPEVSISGYDDNWYLGRTDATLSCDVRSNPE 290

Qy        249 KSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLH----------LTLPQAL--------- 288
               :   :   :       | |      | :| :|          |:  |:
Db        291 PTGYDWSTTSGTFPTSAVAQ------GSQLVIHAVDSLFNTTFVCTVTNAVGMGRAEQVI 344

Qy        289 -----PQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTC--------------- 328
                  | ||:|       | | :   :  ::  |     | |
Db        345 FVRETPNTAGAG-------ATGGIIGGIIAAIIATAVAATGILICRQQRKEQTLQGAEED 397

Qy        329 -EVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLS---DSGQVLLESNI 384
               :: ||       | |   :||:  :| |  :    |     | :  |:|    |  :
Db        398 EDLEGPP------SYKPPTPKAKLEAQEMPSQLFTLGASEHSPLKTPYFDAGASCTEQEM 451

Qy        385 ---KVLPTW---STPVHPRASALPAPPTGSALPDPQTASALPDPPA 424
                 |||    | |:|| |::|     || :| |      | |||
Db        452 PRYHELPTLEERSGPLHPGATSL-----GSPIPVP------PGPPA 486


RESULT 73
SILA_HUMAN
ID   SILA_HUMAN     STANDARD;      PRT;   697 AA.
AC   Q96LC7; Q96G54; Q96LC8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Sialic acid binding Ig-like lectin 10 precursor (Siglec-10) (Siglec-
DE   like protein 2).
GN   SIGLEC10 OR SLG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   MEDLINE=21303047; PubMed=11409878;
RA   Yousef G.M., Ordon M.H., Foussias G., Diamandis E.P.;
RT   "Molecular characterization, tissue expression, and mapping of a novel
RT   Siglec-like gene (SLG2) with three splice variants.";
RL   Biochem. Biophys. Res. Commun. 284:900-910(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Spleen;
RX   MEDLINE=21181584; PubMed=11284738;
RA   Munday J., Kerr S., Ni J., Cornish A.L., Zhang J.Q., Nicoll G.,
RA   Floyd H., Mattei M.-G., Moore P., Liu D., Crocker P.R.;
RT   "Identification, characterization and leucocyte expression of
RT   Siglec-10, a novel human sialic acid-binding receptor.";
RL   Biochem. J. 355:489-497(2001).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Dendritic cell;
RX   MEDLINE=21359381; PubMed=11358961;
RA   Li N., Zhang W., Wan T., Zhang J., Chen T., Yu Y., Wang J., Cao X.;
RT   "Cloning and characterization of Siglec-10, a novel sialic acid
RT   binding member of the Ig superfamily, from human dendritic cells.";
RL   J. Biol. Chem. 276:28106-28112(2001).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 3), ALTERNATIVE SPLICING (ISOFORMS 4
RP   AND 5), MUTAGENESIS OF TYR-667, AND INTERACTION WITH PTPN6.
RX   MEDLINE=22152890; PubMed=12163025;
RA   Kitzig F., Martinez-Barriocanal A., Lopez-Botet M., Sayos J.;
RT   "Cloning of two new splice variants of Siglec-10 and mapping of the
RT   interaction between Siglec-10 and SHP-1.";
RL   Biochem. Biophys. Res. Commun. 296:355-362(2002).
RN   [5]
RP   SEQUENCE OF 334-697 FROM N.A. (ISOFORM 1/3/4).
RC   TISSUE=B-cell;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Putative adhesion molecule that mediates sialic-acid
CC       dependent binding to cells. Preferentially binds to alpha2,3- or
CC       2,6-linked sialic acid. The sialic acid recognition site may be
CC       masked by cis interactions with sialic acids on the same cell
CC       surface. In the immune response, may act as an inhibitory receptor
CC       upon ligand induced tyrosine phosphorylation by recruiting
CC       cytoplasmic phosphatase(s) via their SH2 domain(s) that block
CC       signal transduction through dephosphorylation of signaling
CC       molecules.
CC   -!- SUBUNIT: Interacts with PTPN6/SHP-1 upon phosphorylation.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (isoforms 1-4);
CC       secreted (isoform 5).
CC   -!- ALTERNATIVE PRODUCTS:
```

```
CC       Event=Alternative splicing; Named isoforms=5;
CC       Name=1; Synonyms=Long;
CC         IsoId=Q96LC7-1; Sequence=Displayed;
CC       Name=2; Synonyms=Short, Sv1;
CC         IsoId=Q96LC7-2; Sequence=VSP_002565;
CC       Name=3; Synonyms=Sv3;
CC         IsoId=Q96LC7-3; Sequence=VSP_002564;
CC       Name=4; Synonyms=Sv4;
CC         IsoId=Q96LC7-4; Sequence=VSP_002561;
CC       Name=5; Synonyms=Sv2;
CC         IsoId=Q96LC7-5; Sequence=VSP_002562, VSP_002563;
CC   -!- TISSUE SPECIFICITY: Expressed by peripheral blood leukocytes
CC       (eosinophils, monocytes and a natural killer cell subpopulation).
CC       Isoform 5 is found to be the most abundant isoform. Found in lymph
CC       node, lung, ovary and appendix. Isoform 1 is found at high levels
CC       and isoform 2 at lower levels in bone marrow, spleen and spinal
CC       chord. Isoform 2 is also found in brain. Isoform 4 is specifically
CC       found in natural killer cells.
CC   -!- DOMAIN: Contains 1 copy of a cytoplasmic motif that is referred
CC       to as the immunoreceptor tyrosine-based inhibitor motif (ITIM).
CC       This motif is involved in downmodulation of cellular responses.
CC       The phosphorylated ITIM motif binds to the SH2 domain of
CC       PTPN6/SHP-1.
CC   -!- PTM: Tyrosine phosphorylated.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC
CC       (sialic acid binding Ig-like lectin) family.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AY029277; AAK40255.1; -.
DR   EMBL; AY029277; AAK40256.1; -.
DR   EMBL; AF310233; AAK55139.1; -.
DR   EMBL; AF311905; AAK92542.1; -.
DR   EMBL; AY032685; AAK51124.1; -.
DR   EMBL; BC009955; AAH09955.1; -.
DR   Genew; HGNC:15620; SIGLEC10.
DR   MIM; 606091; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 4.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Cell adhesion; Lectin; Transmembrane; Signal; Glycoprotein;
KW   Immunoglobulin domain; Repeat; Phosphorylation; Alternative splicing;
KW   Polymorphism.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    697       SIALIC ACID BINDING IG-LIKE LECTIN 10.
FT   DOMAIN       17    550       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    551    571       POTENTIAL.
FT   DOMAIN      572    697       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       18    121       IG-LIKE V-TYPE.
FT   DOMAIN      146    231       IG-LIKE C2-TYPE 1.
FT   DOMAIN      251    339       IG-LIKE C2-TYPE 2.
FT   DOMAIN      344    441       IG-LIKE C2-TYPE 3.
FT   SITE        595    600       ITIM MOTIF 1.
FT   SITE        665    670       ITIM MOTIF 2.
FT   DISULFID     36    173       BY SIMILARITY.
FT   DISULFID     41    101       BY SIMILARITY.
FT   DISULFID    164    215       BY SIMILARITY.
FT   DISULFID    276    323       BY SIMILARITY.
FT   DISULFID    380    425       BY SIMILARITY.
FT   MOD_RES     667    667       PHOSPHORYLATION
FT                                (INVOLVED IN BINDING TO PTPN6)
FT                                (PROBABLE).
FT   CARBOHYD    100    100       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    355    355       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    364    364       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    486    486       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    504    504       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    125    214       Missing (in isoform 4).
FT                                /FTId=VSP_002561.
FT   VARSPLIC    140    185       TALTQKPDVYIPETLEPGQPVTVICVFNWAFEECPPPSFSW
FT                                TGAAL -> TGMRWGGNPCLSHWGGTLGTAYGLSREGSQGP
FT                                LQHKNLPPRSLSQP (in isoform 5).
FT                                /FTId=VSP_002562.
FT   VARSPLIC    186    697       Missing (in isoform 5).
FT                                /FTId=VSP_002563.
FT   VARSPLIC    141    198       Missing (in isoform 3).
FT                                /FTId=VSP_002564.
FT   VARSPLIC    445    539       Missing (in isoform 2).
FT                                /FTId=VSP_002565.
FT   MUTAGEN     667    667       Y->F: ABOLISHES BINDING TO PTPN6.
FT   CONFLICT     28     28       S -> P (IN REF. 4).
FT   CONFLICT    134    134       G -> R (IN REF. 4).
FT   CONFLICT    226    226       V -> A (IN REF. 1).
FT   CONFLICT    334    336       RAL -> GTR (IN REF. 5).
FT   CONFLICT    344    344       P -> S (IN REF. 3).
FT   CONFLICT    440    440       L -> P (IN REF. 4).
FT   CONFLICT    587    587       R -> K (IN REF. 3).
FT   CONFLICT    625    625       P -> S (IN REF. 3).
SQ   SEQUENCE   697 AA;  76619 MW;  6CB231CE49411D1B CRC64;

  Query Match             4.6%;  Score 108;  DB 1;  Length 697;
  Best Local Similarity   18.7%;  Pred. No. 6.1;
  Matches   98;  Conservative   53;  Mismatches  166;  Indels  206;  Gaps   22;

Qy          8 RHLLLVLQLALLPA---ATQGNKVVL-GKKGDTVELTCTASQKKSIQFHWKNSNQIKILG 63
              | |:: :     ||     |||   |  :|| : | | |  :      |
Db        239 RDLVISISRDNTPALEPQPQGNVPYLEAQKGQFLRLLCAADSQPPATLSW---------- 288

Qy         64 NQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQ--------- 114
                      | :|  |     |       | :  :|  ||  | |  |::
Db        289 -----------VLQNRVLSSSHPWGPRPLGLELPGVKAGDSGRYTCRAENRLGSQQRALD 337

Qy        115 ------KEEVQLLVF--------GLTANSDTHLLQGQSLTL---TLESPPGSSPSVQCRS 157
                    | ::::|         |   :   :|:|||| |   |  |||
Db        338 LSVQYPPENLRVMVSQANRTVLENLGNGTSLPVLEGQSLCLVCVTHSSPPA--------- 388

Qy        158 PRGKNIQGGKTLSVSQ-----------LELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
               |    | |: || ||           ::::  | :||
Db        389 -RLSWTQRGQVLSPSQPSDPGVLELPRVQVEHEGEFTCHAR------------------- 428

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELW----------WQAERASSSKSWITFDL 256
                  :    : |  |  : ::  || |     |            ||  | | : |:
Db        429 -----HPLGSQHVSLSLSVHYS-PKLLGPSCSWEAEGLHCSCSSQASPAPSLRWWL---- 478

Qy        257 KNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGS-GNLTLALEAKTGKLHQEVNLV 315
              :|:     :||                :  |  ||   | :|:|    | |   :
Db        479 -GEELLEGNSSQDSF-------------EVTPSSAGPWANSSLSLH---GGLSSGL--- 517

Qy        316 VMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVS---------------------- 352
                      | || |     :    |:| :|:  :|
Db        518 ---------RLRCEAWNVHGAQSGSILQLPDKKGLISTAFSNGAFLGIGITALLFLCLAL 568

Qy        353 -------KREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPT----------WSTPVH 395
                    ||         |      :|          | |:||           :||
Db        569 IIMKILPKRRTQTETPRPRFSRHSTIL---------DYINVVPTAGPLAQKRNQKATPNS 619

Qy        396 PRASALPAPPTGSALPDPQTAS---------ALPDPPAASALP 429
              ||    | ||    | |::           : |:| :::  |
Db        620 PRT---PLPP---GAPSPESKKNQKKQYQLPSFPEPKSSTQAP 656


RESULT 74
```

```
TIE1_HUMAN
ID   TIE1_HUMAN     STANDARD;      PRT;  1138 AA.
AC   P35590;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase receptor Tie-1 precursor (EC 2.7.1.112).
GN   TIE OR TIE1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92195316; PubMed=1312667;
RA   Partanen J., Armstrong E., Maekelae T.P., Korhonen J., Sandberg M.,
RA   Renkonen R., Knuutila S., Huebner K., Alitalo K.;
RT   "A novel endothelial cell surface receptor tyrosine kinase with
RT   extracellular epidermal growth factor homology domains.";
RL   Mol. Cell. Biol. 12:1698-1707(1992).
RN   [2]
RP   REVISIONS.
RA   Partanen J.M.;
RL   Submitted (JUL-1993) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Probable protein tyrosine-kinase transmembrane receptor.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: Specifically expressed in developing vascular
CC       endothelial cells.
CC   -!- SIMILARITY: Belongs to the Tyr family of protein kinases. Tie
CC       subfamily.
CC   -!- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 3 fibronectin type III domains.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X60957; CAA43290.1; -.
DR   HSSP; P11362; 1FGK.
DR   Genew; HGNC:11809; TIE.
DR   MIM; 600222; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004714; F:transmembrane receptor protein tyrosine kin. . .; TAS.
DR   GO; GO:0007498; P:mesoderm development; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF00047; ig; 2.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00060; FN3; 3.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS50026; EGF_3; 2.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   Receptor; Tyrosine-protein kinase; Transferase; Signal; ATP-binding;
KW   Repeat; EGF-like domain; Transmembrane; Immunoglobulin domain;
KW   Glycoprotein; Phosphorylation; Multigene family.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1138       TYROSINE-PROTEIN KINASE RECEPTOR TIE-1.
FT   DOMAIN       25    759       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    760    784       POTENTIAL.
FT   DOMAIN      785   1138       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       43    105       IG-LIKE C2-TYPE 1.
FT   DOMAIN      214    256       EGF-LIKE 1.
FT   DOMAIN      258    303       EGF-LIKE 2.
FT   DOMAIN      305    345       EGF-LIKE 3.
FT   DOMAIN      372    426       IG-LIKE C2-TYPE 2.
FT   DOMAIN      446    540       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      543    639       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      643    744       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      839   1118       PROTEIN KINASE.
FT   NP_BIND     845    853       ATP (BY SIMILARITY).
FT   BINDING     870    870       ATP (BY SIMILARITY).
FT   ACT_SITE    979    979       BY SIMILARITY.
FT   CARBOHYD     83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    503    503       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    596    596       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    709    709       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   MOD_RES    1007   1007       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   1138 AA;  125089 MW;  3B42BE33678C58A1 CRC64;

  Query Match             4.6%;  Score 107.5;  DB 1;  Length 1138;
  Best Local Similarity   20.9%;  Pred. No. 12;
  Matches   83;  Conservative   44;  Mismatches  136;  Indels  135;  Gaps   18;

Qy        109 CEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPP------GSSP-----SVQCRS 157
              ||  |:  ::                  |     |   ||: |        :|     |:: |
Db        344 CEKSDRIPQI-------------LNMASELEFNLETMPRINCAAAGNPFPVRGSIELRK 389

Qy        158 PRGKNIQGGKTL----------SVSQLELQDSGTWTCTVLQN--QKKVEFKIDIVVLAFQ 205
              | |  :   | :          | :| | ||| | | |  :  |    ||::: |
Db        390 PDGTVLLSTKAIVEPEKTTAEFEVPRLVLADSGFWECRVSTSGGQDSRRFKVNVKVPPVP 449

Qy        206 KASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT-----------F 254
              |:  :  |:  |:  |  ::|: :    :  | :: :   |:  | |
Db        450 LAAPRLLTKQSRQLVVSPLVSFSGDGPISTVRLHYRPQ--DSTMDWSTIVVDPSENVTLM 507

Qy        255 DLKNK-----EVSVKR-------------VTQD---PKLQM---------GKKLPLHLT 283
              :|: |      | : |             :| |   | ||           :| :  :
Db        508 NLRPKTGYSVRVQLSRPGEGGEGAWGPPTLMTTDCPEPLLQPWLEGWHVEGTDRLRVSWS 567

Qy        284 LPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLT-------------CEV 330
              ||       | | |    :  |:  :|       | |   ||              | :
Db        568 LPLVPGPLVGDGFLLRLWDGTRGQERRENVSSPQARTALLTGLTPGTHYQLDVQLYHCTL 627

Qy        331 WGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAG------MWQCLLSDSGQVLLESNI 384
              || ||                     |  || | :|      :    ||||
Db        628 LGPASP-------------------PAHVLLPPSGPPAPRHLHAQALSDS--------- 658

Qy        385 KVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDP 422
              ::  ||    ||   |||  |    : : | |    ||
Db        659 EIQLTWK---HP--EALPG-PISKYVVEVQVAGGAGDP 690


RESULT 75
NEO1_CHICK
ID   NEO1_CHICK     STANDARD;      PRT;  1443 AA.
AC   Q90610;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
```

```
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neogenin (Fragment).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Embryonic brain;
RX   MEDLINE=95105243; PubMed=7806578;
RA   Vielmetter J., Roman J.M., Dreyer W.J.;
RT   "Neogenin, an avian cell surface protein expressed during terminal
RT   neuronal differentiation, is closely related to the human tumor
RT   suppressor molecule deleted in colorectal cancer.";
RL   J. Cell Biol. 127:2009-2020(1994).
CC   -!- FUNCTION: May be involved as a regulatory protein in the
CC       transition of undifferentiated proliferating cells to their
CC       differentiated state. May also function as a cell adhesion
CC       molecule in a broad spectrum of embryonic and adult tissues.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DEVELOPMENTAL STAGE: In retina, expressed on ganglion cell fibers
CC       as soon as they begin to extend their axons.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC
CC       family.
CC   -!- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 6 fibronectin type III domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U07644; AAC59662.1; -.
DR   PIR; I50600; I50600.
DR   HSSP; P11276; 2MFN.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 4.
KW   Cell adhesion; Repeat; Transmembrane; Immunoglobulin domain;
KW   Glycoprotein.
FT   NON_TER       1      1
FT   DOMAIN       <1   1090       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1091   1111       POTENTIAL.
FT   DOMAIN     1112   1443       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       18    113       IG-LIKE C2-TYPE 1.
FT   DOMAIN      118    204       IG-LIKE C2-TYPE 2.
FT   DOMAIN      212    302       IG-LIKE C2-TYPE 3.
FT   DOMAIN      307    392       IG-LIKE C2-TYPE 4.
FT   DOMAIN      422    519       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      522    615       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      616    714       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      720    814       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      835    935       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      936   1037       FIBRONECTIN TYPE-III 6.
FT   DISULFID     40     95       BY SIMILARITY.
FT   DISULFID    139    187       BY SIMILARITY.
FT   DISULFID    236    286       BY SIMILARITY.
FT   DISULFID    328    376       BY SIMILARITY.
FT   CARBOHYD     39     39       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    176    176       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    456    456       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    475    475       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    625    625       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    700    700       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    894    894       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1443 AA;  158050 MW;  558C6795579C0E26 CRC64;

  Query Match             4.6%;  Score 107.5;  DB 1;  Length 1443;
  Best Local Similarity   19.3%;  Pred. No. 17;
  Matches   95;  Conservative   68;  Mismatches  199;  Indels  129;  Gaps   21;

Qy         14 LQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGP 73
              | :| ||  |   ::    ||::  | |  :   :    |:   |               |
Db        112 LTVAGLPRFTSQPELSSVYKGNSAILNCEVNVDLAPFVRWEQDRQ-------------P 157

Qy         74 SKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED---------------QKEE 117
               |:||      ::   :  |:| |    |   | | :|                    ||
Db        158 LSLDDR------VFKLPSGALLIGNATDTDGGFYRCVIESGGTPKYSEEAELKILPDPEE 211

Qy        118 VQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKN----------IQGG 166
               | |||    :| |  : ||:      :  |  |:   | :  |:          :: |
Db        212 PQSLVFVRQPSSLTK-VTGQNAVFPCVA--GGFPTPYVRWTKNGEELITEDSERFALRAG 268

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA---FQKASSIVYKKEGEQVEFSF 223
              :| :| :  :| ||:||      : :| : :: |     | |  : :|  |   : |
Db        269 GSLLISDVTEEDVGTYTCIADNENETIEAQAELAVQVPPEFLKRPANIYAHESMDIVFEC 328

Qy        224 PLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVK----RVTQDPKLQM----- 274
                  ::||          : : :  |:    || :| :     :: ::  ||:
Db        329 -------EVTG---------KPTPTVKWV----KNGDVVIPSDYFKIVKEHNLQVLGLVK 368

Qy        275 -------------------GKK---LPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEV 312
                              | :   | | : :|   |   |      |  | | | |
Db        369 SDEGFYQCIAENDVGNAQAGAQLIILDLDVAIPTLPPTSLTSATNDHLAPATTGPLPTAP 428

Qy        313 NLVVMRATQLQKNLTCEVW-----GPTSPKLMLSLKLENKEA----KVSKREKPVWVLNP 363
              ||  || :        |      |    |  |:    ||     :|    :|
Db        429 RDVV--ATLVSTRFIRLTWRTPVSDPQGDNLTYSI-FYTKEGINRERVENTSRP------ 479

Qy        364 EAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPP 423
              |  | ::   | |:  : |    :   |    :  ||   :  |: |     |:
Db        480 --GETQVMI----QNLMPETVYVFRVVAQNKHGHGES-SAPLKVATQPEVQLPGPAPNIR 532

Qy        424 AASALPAALAV 434
              | :  | :: |
Db        533 AYAGSPTSVTV 543


RESULT 76
CSA_DICDI
ID   CSA_DICDI      STANDARD;      PRT;   514 AA.
AC   P08796; P19408;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Contact site A protein precursor (CSA) (Membrane-associated
DE   glycoprotein GP80) (Cell adhesion molecule GP80).
GN   CSAA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Noegel A., Gerisch G., Stadler J., Westphal M.;
RT   "Complete sequence and transcript regulation of a cell adhesion
RT   protein from aggregating Dictyostelium cells.";
RL   EMBO J. 5:1473-1476(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89105362; PubMed=3063296;
RA   Siu C.-H., Wong L.M., Lam T.Y., Kamboj R.K., Choi A., Cho A.;
RT   "Molecular mechanisms of cell-cell interaction in Dictyostelium
RT   discoideum.";
```

```
RL   Biochem. Cell Biol. 66:1089-1099(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX2;
RX   MEDLINE=92406924; PubMed=1326559;
RA   Desbarats L., Lam T.Y., Wong L.M., Siu C.H.;
RT   "Identification of a unique cAMP-response element in the gene
RT   encoding the cell adhesion molecule gp80 in Dictyostelium
RT   discoideum.";
RL   J. Biol. Chem. 267:19655-19664(1992).
RN   [4]
RP   SEQUENCE OF 20-49.
RA   Wong L.M., Siu C.-H.;
RT   "Cloning of cDNA for the contact site A glycoprotein of Dictyostelium
RT   discoideum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:4248-4252(1986).
RN   [5]
RP   CELL-BINDING DOMAIN.
RX   MEDLINE=89034443; PubMed=3182938;
RA   Kamboj R.K., Wong L.M., Lam T.Y., Siu C.H.;
RT   "Mapping of a cell-binding domain in the cell adhesion molecule gp80
RT   of Dictyostelium discoideum.";
RL   J. Cell Biol. 107:1835-1843(1988).
RN   [6]
RP   GPI-ANCHOR.
RX   MEDLINE=89251561; PubMed=2721485;
RA   Stadler J., Keenan T.W., Bauer G., Gerisch G.;
RT   "The contact site A glycoprotein of Dictyostelium discoideum carries
RT   a phospholipid anchor of a novel type.";
RL   EMBO J. 8:371-377(1989).
CC   -!- FUNCTION: This cell-surface glycoprotein mediates cell-cell
CC       binding via homophilic interaction.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       that contains a phosphoceramide moiety. Such anchor mediates a
CC       fast and long persistence cell adhesion of the protein.
CC   -!- DEVELOPMENTAL STAGE: Restricted to the aggregation stage of
CC       development of D.discoideum.
CC   -!- PTM: Acylated, phosphorylated on serine and N-glycosylated with
CC       two types of oligosaccharide chains.
CC   -!- MISCELLANEOUS: The expression of this stringently regulated
CC       protein during cell development is mediated through cell-surface
CC       cAMP receptors.
CC   -!- SIMILARITY: THE C-TERMINAL REGION CONTAINS CLUSTERS OF PROLINE
CC       REGULARLY ALTERNATING WITH A HYDROXYAMINO ACID AND SHOWS
CC       SIMILARITY TO THE HINGE REGION OF IMMUNOGLOBULINS. THIS DOMAIN
CC       MIGHT ACT AS A SPACER TO ELEVATE SITES ACTIVE IN CELL CONTACT
CC       INTO THE EXTRACELLULAR SPACE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04004; CAA27634.1; -.
DR   EMBL; M36545; AAA33212.1; -.
DR   EMBL; X66483; CAA47110.1; -.
DR   PIR; S22066; A31643.
DR   DictyBase; DDB0001860; csaA.
DR   InterPro; IPR002909; IPT_TIG.
DR   Pfam; PF01833; TIG; 2.
KW   Cell adhesion; Glycoprotein; Membrane; Phosphorylation; Repeat;
KW   GPI-anchor; Lipoprotein; Signal.
FT   SIGNAL       1     19
FT   CHAIN       20    492       CONTACT SITE A PROTEIN.
FT   PROPEP     493    514       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN      20    453       GLOBULAR (POTENTIAL).
FT   DOMAIN     454    485       PRO-RICH (HINGE STRUCTURE).
FT   DOMAIN     462    479       2 X 8 AA REPEATS, PRO-RICH.
FT   REPEAT     462    469       1.
FT   REPEAT     472    479       2.
FT   CARBOHYD   128    128       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   137    137       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   207    207       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   294    294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   399    399       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID      492    492       GPI-like-anchor amidated serine
FT                               (Potential).
FT   CONFLICT   216    216       V -> G (IN REF. 1).
SQ   SEQUENCE   514 AA;  53696 MW;  96A1C7CF42FEC096 CRC64;

  Query Match             4.6%;  Score 107;  DB 1;  Length 514;
  Best Local Similarity   22.4%;  Pred. No. 4.8;
  Matches  102;  Conservative   52;  Mismatches  164;  Indels  138;  Gaps   24;

Qy         63 GNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDS-DTYICEVEDQKEEVQLL 121
              |  |  |  ||| ::                    | : || :: :  |    : |
Db        116 GRIGMILVDGPSNISG-----------------YKLNVNDSINSAMLSV--TADSVSPT 155

Qy        122 VFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRS----------------PRGKNIQ 164
              :: |  |:    : |  | | |  | | |  | :|                |    :
Db        156 IYFLVPNT----IAGGLLNLELIQPFGFSTIVTSKSVFSPTITSITPLAFDLTPTNVTVT 211

Qy        165 GG---KTLSVSQ-------LELQDSGTWTCTVLQNQKKV-EFKIDIVVLAFQKASSIVYK 213
              |      | ||:        | :|| ||  | |:   : | |    |   |       |
Db        212 GKYFVTTASVTMGSHIYTGLTVQDDGT-NCHVIFTTRSVYESSNTITAKASTGVDMIYLD 270

Qy        214 KEGEQVEFSF----PLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQD 269
              :| |  :|    |  : :::  | |:       |:: :   |     ::|: |
Db        271 NQGNQQPITFTYNPPTITSTKQVNDSVEI-------STTNTGTDF----TQISLTMGTSS 319

Qy        270 PK--LQMGKKLPLHLTLPQALPQ-------YAGSGN-------LTLALEAKTGKLHQ--- 310
              |   :  |    : :||| |||:         ||  |       :|  : : |   |
Db        320 PTNLVITGTNEKIVITLPHALPEGEIQFNLKAGISNVVTSTLLVTPVINSVTQAPHNGGS 379

Qy        311 ----------EVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVW 359
                       |::||       :| |  |  | |           | |: :   |
Db        380 ITISGIFLNNAHVSIVV------DQNTTDIVCAPDS----------NGESIICPVEAGSG 423

Qy        360 VLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV------HPRASALPAP-PTGSALPD 412
              :|    ::   ||        ||   | ||          | :| |:| || :| |
Db        424 TINLVVTNYKNFASD-------PTIKTEATTSTTYTIPDTPTPTDTATPSPTPTETATPS 476

Qy        413 P-----QTASALPDPPAASAL--PAALAV--ISFLL 439
              |      |       | :|: |  | :| |  |||:|
Db        477 PTPKPTSTPEETEAPSSATTLISPLSLIVIFISFVL 512


RESULT 77
CEA1_MOUSE
ID   CEA1_MOUSE     STANDARD;      PRT;   521 AA.
AC   P31809;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Carcinoembryonic antigen-related cell adhesion molecule 1 precursor
DE   (Biliary glycoprotein 1) (BGP-1) (Murine hepatitis virus receptor)
DE   (MHV-R) (Biliary glycoprotein D).
GN   CEACAM1 OR BGP OR BGP1 OR BGPD.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD-1; TISSUE=Colon;
RX   MEDLINE=93100785; PubMed=8380065;
RA   Dveksler G.S., Dieffenback C.B., Cardellichio C.B., McCuaig K.,
RA   Pensiero M.N.;
RT   "Several members of the mouse carcinoembryonic antigen-related
RT   glycoprotein family are functional receptors for the coronavirus
RT   mouse hepatitis virus-A59.";
RL   J. Virol. 67:1-8(1993).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=92046352; PubMed=1719235;
RA   Dveksler G.S., Pensiero M.N., Cardellichio C.B., Williams R.K.,
RA   Jiang G.-S., Holmes K.V., Dieffenbach C.W.;
RT   "Cloning of the mouse hepatitis virus (MHV) receptor: expression in
RT   human and hamster cell lines confers susceptibility to MHV.";
RL   J. Virol. 65:6881-6891(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93273228; PubMed=8500759;
RA   McCuaig K., Rosenberg M., Nedellec P., Turbide C., Beauchemin N.;
RT   "Expression of the Bgp gene and characterization of mouse colon
RT   biliary glycoprotein isoforms.";
RL   Gene 127:173-183(1993).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=CD-1; TISSUE=Colon;
RX   MEDLINE=89195121; PubMed=2702644;
RA   Beauchemin N., Turbide C., Afar D., Raymond M., Bell J.,
RA   Stanners C.P., Fuks A.;
RT   "A mouse analogue of the human carcinoembryonic antigen.";
RL   Cancer Res. 49:2017-2021(1989).
RN   [5]
RP   SEQUENCE OF 35-59.
RX   MEDLINE=91288498; PubMed=1648219;
RA   Williams R.K., Jiang G.-S., Holmes K.V.;
RT   "Receptor for mouse hepatitis virus is a member of the
RT   carcinoembryonic antigen family of glycoproteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:5533-5536(1991).
CC   -!- FUNCTION: Unknown; receptor for murine coronavirus MHV-A59.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=P31809-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=P31809-2; Sequence=VSP_002484, VSP_002485;
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67279; CAA47696.1; -.
DR   EMBL; M77196; AAA37858.1; -.
DR   EMBL; X15351; CAA33409.1; -.
DR   PIR; JC1505; WMMSR1.
DR   PIR; JC1508; JC1508.
DR   MGD; MGI:1347245; Ceacam1.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 3.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain; Glycoprotein; Transmembrane; Signal;
KW   Repeat; Alternative splicing; Receptor.
FT   SIGNAL       1     34
FT   CHAIN       35    521       CARCINOEMBRYONIC ANTIGEN-RELATED CELL
FT                               ADHESION MOLECULE 1.
FT   DOMAIN       35    428       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    429    447       POTENTIAL.
FT   DOMAIN      448    521       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       35    142       IG-LIKE V-TYPE.
FT   DOMAIN      147    234       IG-LIKE C2-TYPE 1.
FT   DOMAIN      239    319       IG-LIKE C2-TYPE 2.
FT   DOMAIN      323    411       IG-LIKE C2-TYPE 3.
FT   DISULFID    167    217       PROBABLE.
FT   DISULFID    261    301       PROBABLE.
FT   DISULFID    346    394       PROBABLE.
FT   CARBOHYD     71     71       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     89     89       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    199    199       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    206    206       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    210    210       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    226    226       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    258    258       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    290    290       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    294    294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    304    304       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    317    317       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    375    375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    455    458       GSDQ -> SGSF (in isoform Short).
FT                               /FTId=VSP_002484.
FT   VARSPLIC    459    521       Missing (in isoform Short).
FT                               /FTId=VSP_002485.
SQ   SEQUENCE   521 AA;  57015 MW;  1C8F71FAC47DD54E CRC64;

  Query Match             4.6%;  Score 107;  DB 1;  Length 521;
  Best Local Similarity   18.4%;  Pred. No. 4.9;
  Matches   80;  Conservative   63;  Mismatches  142;  Indels  150;  Gaps   18;

Qy         16 LALLPAATQGNKVVLGKKGDTVELTCTA-SQKKSIQFHWKNSNQIKILGNQGSFLTKGPS 74
              :| | |  | |   : :  |:| ||| : :   :| : |  :          |  |::|
Db        143 ILLKPNITSNNSNPV-EGDDSVSLTCDSYTDPDNINYLWSRN---------GESLSEGD- 191

Qy         75 KLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVE-----DQKEEVQL-LVFG---- 124
                        |   :||  | : |:    |:  |:||       :: :   | :::|
Db        192 ---------RLKLSEGNRTLTLLNVTRNDTGPYVCETRNPVSVNRSDPFSLNIIYGPDTP 242

Qy        125 LTANSDTHLLQGQSLTLTLESPPGSSPSVQC-----RSPRGKNIQGGKTLSVSQLELQDS 179
              : : || :|  | :| |:  :   |:|  |        |   :    : | :  :   :|
Db        243 IISPSDIYLHPGSNLNLSCHA--ASNPPAQYFWLINEKPHASS----QELFIPNITTNNS 296

Qy        180 GTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELW 239
              ||:||                                          |    :||
Db        297 GTYTC-------------------------------------------FVNNSVTG----- 309

Qy        240 WQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTL 299
                    |:: :    | ::|     | ||:
Db        310 -------LSRTTV------KNITVLEPVTQPFLQV------------------------- 331

Qy        300 ALEAKTGKLHQEVNLVVMRATQLQKNLTCEV-WGPTSPKLMLSLKLENKEAKVSKREKPV 358
                  |    :|:: | :  | |  ::   : |   |  | |: ::   :     | |
Db        332 -----TNTTVKELDSVTL--TCLSNDIGANIQWLFNSQSLQLTERMTLSQNNSILRIDP- 383

Qy        359 WVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASA 418
              :   :|| :|| :|:   |   ::||:              :   ||  | |   |
Db        384 -IKREDAGEYQCEISNPVSVRRSNSIKL------------DIIFDPTQGGLSDGAIAGI 429

Qy        419 LPDPPAASALPAALA 433
              :    |  || | ||
Db        430 VIGVVAGVALIAGLA 444


RESULT 78
SCA4_RICPR
ID   SCA4_RICPR     STANDARD;      PRT;  1022 AA.
AC   Q9ZD49; Q9AJ36; Q9ZD48;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE   (Protein PS 120).
GN   SCA4 OR RP498/RP499.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
```

```
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
RN   [2]
RP   SEQUENCE OF 11-1016 FROM N.A.
RA   Sekeyova Z., Roux V., Raoult D.;
RT   "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RT   'gene D' coding for an intracytoplasmic protein.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 234.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235272; CAA14951.1; ALT_FRAME.
DR   EMBL; AJ235272; CAA14950.1; ALT_FRAME.
DR   EMBL; AF200340; AAK31305.1; -.
KW   Antigen; Complete proteome.
FT   CONFLICT     11     15       EFDPL -> RPGLV (IN REF. 2).
FT   CONFLICT    365    365       H -> Y (IN REF. 2).
FT   CONFLICT    413    413       MISSING (IN REF. 2).
FT   CONFLICT    957    957       G -> R (IN REF. 2).
SQ   SEQUENCE   1022 AA;  114410 MW;  03230E3A663A9622 CRC64;

  Query Match             4.6%;  Score 106.5;  DB 1;  Length 1022;
  Best Local Similarity   21.5%;  Pred. No. 13;
  Matches   83;  Conservative   63;  Mismatches  155;  Indels   85;  Gaps   19;

Qy         64 NQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVF 123
              |    | :  :|  |     |  | :| |: ||  |   ||: |     :   |: ::: :
Db         98 NTNPELAEQIAKEEDDRKFRAFLSNQDNYALI--NKAFEDTKT-----KKNLEKAEIVGY 150

Qy        124 G--LTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLE---LQ 177
               |:  |  :  ||           |  |           :|    :| | |
Db        151 KNVLSTYSVANGYQG----------GFQP----------VQWENQVSASDLRSTVVKN 188

Qy        178 DSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQV------EFSFPLAFTVEK 231
              | |   ||:  |:  |: |             || |::| ||      | :||:   ::|
Db        189 DEGEELCTL--NETTVKTK----------DLIVAKQDGTQVQINSYREINFPI--KLDK 233

Qy        232 LTGSGELWWQAERASSSKSWITFDLKNKEV-SVKRVTQDPKLQMGKKLPLHLTLPQALPQ 290
               ||  |   | :|  :|       |:| |       : |    ||     :: || | :
Db        234 ANGSMHLSMVALKADGTK-----PAKDKAVYFTAHYEEGPN---GKPQLKEISSPQPL-K 284

Qy        291 YAGSGNLTLALEAKTGKLHQEVNLVVMRA--TQLQKNLTCEVWGPTSPKLMLSLKLENKE 348
              : |:|:  :|     |:::    | | |    :: | :       :    ::  | : :
Db        285 FVGTGDDAVAYIEHGGEIY---TLAVTRGKYKEMMKEVALNHGQSVALSQTIAEDLTHVQ 341

Qy        349 AKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGS 408
               :  ||: : | |       | | :      :  : |::  : |: | :
Db        342 GPSHETHKPIIIPNQE-------LESSIEQHTSQQVPPITTFNKSLQPKISQI-----HQ 389

Qy        409 ALPDPQTASALPDP--PAASALPAAL 432
               |    :| :|:|   ||:||  ::
Db        390 LQPQQAQSSGIPNPVLNAANALSTSM 415

RESULT 79
```

```
OPCM_BOVIN
ID   OPCM_BOVIN     STANDARD;      PRT;   345 AA.
AC   P11834;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Opioid binding protein/cell adhesion molecule precursor (OBCAM)
DE   (Opioid-binding cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM OR OCAM.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=89251576; PubMed=2721489;
RA   Schofield P.R., McFarland K.C., Hayflick J.S., Wilcox J.N.,
RA   Cho T.M., Roy S., Lee N.M., Loh H.H., Seeburg P.H.;
RT   "Molecular characterization of a new immunoglobulin superfamily
RT   protein with potential roles in opioid binding and cell contact.";
RL   EMBO J. 8:489-495(1989).
CC   -!- FUNCTION: Binds opioids in the presence of acidic lipids; probably
CC       involved in cell contact.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC       family.
CC   -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12672; CAA31192.1; -.
DR   PIR; S03199; S03199.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal; Lipoprotein.
FT   SIGNAL        1     27
FT   CHAIN        28    322       OPIOID BINDING PROTEIN/CELL ADHESION
FT                                 MOLECULE.
FT   PROPEP      323    345       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       39    126       IG-LIKE C2-TYPE 1.
FT   DOMAIN      136    219       IG-LIKE C2-TYPE 2.
FT   DOMAIN      223    310       IG-LIKE C2-TYPE 3.
FT   DISULFID     57    115       POTENTIAL.
FT   DISULFID    157    202       POTENTIAL.
FT   DISULFID    244    296       POTENTIAL.
FT   CARBOHYD     44     44       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   CARBOHYD     70     70       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (PROBABLE).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       322    322       GPI-anchor amidated asparagine
FT                                 (Potential).
SQ   SEQUENCE   345 AA;  37914 MW;  D1ECC8D9E7D8CB19 CRC64;

  Query Match             4.6%;  Score 106;  DB 1;  Length 345;
  Best Local Similarity   21.3%;  Pred. No. 3.4;
  Matches   48;  Conservative   40;  Mismatches   83;  Indels   54;  Gaps    8;

Qy         11 LLVLQLALLPAATQGNKVVLG------------KKGDTVELTCTASQKKSIQFHWKNSN 57
```

```
Db         14 LVVVSLRLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVT-RVAWLNRS 72

Qy         58 QIKILGN-------QGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICE 110

Db         73 TILYAGNDKWSIDPRVIILVNTPTQ---------------YSIMIQNVDVYDEGPYTCS 116

Qy        111 VE----DQKEEVQLLVFG----LTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKN 162

Db        117 VQTDNHPKTSRVHLIVQVPPQIMNISSDVTVNEGSSVTLLCLAIGRPEPTVTWRHLSVKE 176

Qy        163 IQG----GKTLSVSQLELQDSGTWTCTVLQ-----NQKKVEFKID 198

Db        177 GQGFVSEDEYLEISDIKRDQSGEYECSALNDVAAPDVRKVKITVN 221


RESULT 80
PERT_BORBR
ID   PERT_BORBR     STANDARD;      PRT;   916 AA.
AC   Q03035;
DT   01-JUL-1993 (Rel. 26, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Pertactin precursor (Outer membrane protein P.68) (P.94).
GN   PRN OR BB1366.
OS   Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=518;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 38-47.
RC   STRAIN=CN7531;
RX   MEDLINE=92407514; PubMed=1527510;
RA   Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT   "Cloning, nucleotide sequence and heterologous expression of the
RT   protective outer-membrane protein P.68 pertactin from Bordetella
RT   bronchiseptica.";
RL   J. Gen. Microbiol. 138:1697-1705(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RB50 / ATCC BAA-588;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
CC   -!- FUNCTION: Agglutinogen that binds to eukaryotic cells; a process
CC       mediated by the R-G-D sequence. Pertactin may have a role in
CC       bacterial adhesion, and thus play a role in virulence.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- DISEASE: May contribute to the disease state of whooping cough.
CC   -!- MISCELLANEOUS: Synthesized only in the presence of low Mg(2+)
CC       concentrations.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54815; CAA38584.1; -.
DR   EMBL; A19180; CAA01453.1; -.
DR   EMBL; BX640441; CAE31864.1; -.
DR   PIR; A47675; A47675.
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   InterPro; IPR004899; Pertactin.
DR   InterPro; IPR003991; Pertactin_C.
DR   InterPro; IPR003992; Pertactin_N.
DR   Pfam; PF03797; Autotransporter; 1.
DR   Pfam; PF03212; Pertactin; 1.
DR   PRINTS; PR01482; PERTACTIN.
DR   PRINTS; PR01484; PRTACTNFAMLY.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 1.
KW   Outer membrane; Signal; Virulence; Repeat; Complete proteome.
FT   SIGNAL        1     37
FT   CHAIN        38    916       P.94.
FT   CHAIN        38    717       PERTACTIN (P.68).
FT   PROPEP      718    916       POTENTIAL.
FT   SITE        263    265       CELL ATTACHMENT SITE (INVOLVED IN
FT                                ADHESION TO VARIOUS EUKARYOTIC CELL
FT                                LINES).
FT   SITE        706    708       CELL ATTACHMENT SITE (POTENTIAL).
FT   DOMAIN      269    288       4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT   REPEAT      269    273       1.
FT   REPEAT      274    278       2.
FT   REPEAT      279    283       3.
FT   REPEAT      284    288       4 (APPROXIMATE).
FT   DOMAIN      578    606       6 X 3 AA REPEATS OF P-Q-P.
FT   CONFLICT     10     10       K -> L (IN REF. 1).
FT   CONFLICT     29     31       MISSING (IN REF. 1).
FT   CONFLICT     36     36       H -> Y (IN REF. 1).
FT   CONFLICT    281    285       MISSING (IN REF. 1).
FT   CONFLICT    471    471       V -> C (IN REF. 1).
FT   CONFLICT    510    510       W -> L (IN REF. 1).
FT   CONFLICT    521    521       A -> G (IN REF. 1).
FT   CONFLICT    595    595       P -> PQPP (IN REF. 1).
FT   CONFLICT    803    803       A -> S (IN REF. 1).
SQ   SEQUENCE   916 AA;  94352 MW;  BA9C62D4E94E49AA CRC64;

  Query Match             4.5%;  Score 105.5;  DB 1;  Length 916;
  Best Local Similarity   24.9%;  Pred. No. 13;
  Matches   58;  Conservative   19;  Mismatches  101;  Indels   55;  Gaps    9;

Qy        239 WWQAERASSSKS-----WITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAG 293

Db        422 WTGATRAVDSLSIDNATWVMTD--NSNVGALRLASDGSVDFQQPAEAGRFKVLMVDTLAG 479

Qy        294 SGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWG--PTSPKLML----------S 341

Db        480 SGLFRMNVFADLGLSDK---LVVMRDASGQHRLWVRNSGSEPASANTMLLVQTPRGSAAT 536

Qy        342 LKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWS-----TPVHP 396

Db        537 FTLANKDGKV------------DIGTYRYRLAANGN----------GQWSLVGAKAPPAP 574

Qy        397 RASALPAPPTGSALPDP------QTASALPDPPAASALPAALAVISFLLGLGL 443

Db        575 KPAPQPGPQPGPQPPQPPQPPQRQPEAPAPQPPAGRELSAAANAAVNTGGVGL 627


RESULT 81
2ACA_HUMAN
ID   2ACA_HUMAN     STANDARD;      PRT;  1150 AA.
AC   Q06190; Q06189; Q9NPQ5;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Serine/threonine protein phosphatase 2A, 72/130 kDa regulatory subunit
DE   B (PP2A, subunit B, B''-PR72/PR130) (PP2A, subunit B, B72/B130
DE   isoforms) (PP2A, subunit B, PR72/PR130 isoforms) (PP2A, subunit B, R3
DE   isoform).
GN   PPP2R3A OR PPP2R3.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC   TISSUE=Fetal brain;
RX   MEDLINE=93315512; PubMed=8392071;
RA   Hendrix P., Mayer-Jaekel R.E., Cron P., Goris J., Hofsteenge J.,
RA   Merlevede W., Hemmings B.A.;
RT   "Structure and expression of a 72-kDa regulatory subunit of protein
RT   phosphatase 2A. Evidence for different size forms produced by
RT   alternative splicing.";
RL   J. Biol. Chem. 268:15267-15276(1993).
RN   [2]
RP   SEQUENCE OF 768-1150 FROM N.A.
RA   Pluvinet R., Estivill X., Escarceller M., Sumoy L.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: The B regulatory subunit might modulate substrate
CC       selectivity and catalytic activity, and also might direct the
CC       localization of the catalytic enzyme to a particular subcellular
CC       compartment.
CC   -!- SUBUNIT: PP2A consists of a common heterodimeric core enzyme,
CC       composed of a 36 kDa catalytic subunit (subunit C) and a 65 kDa
CC       constant regulatory subunit (PR65 or subunit A), that associates
CC       with a variety of regulatory subunits. Proteins that associate
CC       with the core dimer include three families of regulatory subunits
CC       B (the R2/B/PR55/B55, R3/B''/PR72/PR130/PR59 and R5/B'/B56
CC       families), the 48 kDa variable regulatory subunit, viral proteins,
CC       and cell signaling molecules.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=PR130; Synonyms=130 kDa;
CC         IsoId=Q06190-1; Sequence=Displayed;
CC       Name=PR72; Synonyms=72 kDa;
CC         IsoId=Q06190-2; Sequence=VSP_005107, VSP_005108;
CC   -!- TISSUE SPECIFICITY: Expressed in heart, brain, placenta, lung,
CC       muscle and kidney.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L07590; AAB02613.1; -.
DR   EMBL; L12146; AAB02614.1; -.
DR   EMBL; AL389975; CAB97532.1; -.
DR   PIR; A47114; A47114.
DR   PIR; B47114; B47114.
DR   Genew; HGNC:9307; PPP2R3A.
DR   MIM; 604944; -.
DR   GO; GO:0008601; F:protein phosphatase type 2A, intrinsic regu. . .; TAS.
DR   InterPro; IPR002048; EF-hand.
DR   Pfam; PF00036; efhand; 1.
DR   PROSITE; PS00018; EF_HAND; 1.
KW   Calcium-binding; Multigene family; Alternative splicing.
FT   DOMAIN      670    695       PRO-RICH.
FT   CA_BIND     985    996       EF-HAND (POTENTIAL).
FT   VARSPLIC      1    621       Missing (in isoform PR72).
FT                               /FTId=VSP_005107.
FT   VARSPLIC    622    665       MQILQETLTTSSQANLSVCRSPVGDKAKDTTSAVLIQQTPE
FT                               VIK -> MMIKETSLRRDPDLRGELAFLARGCDFVLPSRFK
FT                               KRLKSFQQTQ (in isoform PR72).
FT                               /FTId=VSP_005108.
SQ   SEQUENCE   1150 AA;  130277 MW;  97A31BA4206518A3 CRC64;

  Query Match             4.5%;  Score 105.5;  DB 1;  Length 1150;
  Best Local Similarity  19.9%;  Pred. No. 17;
  Matches   99;  Conservative   85;  Mismatches  205;  Indels  109;  Gaps   24;

Qy         25 GNKVVLGKKGDTVE---------LTCTASQKKSIQFHWKNSNQ------IKILGNQGS-- 67
               |: :  |   ||:          :|  ||  | : | :  |          ::: |  |
Db        259 GSSISEGSGNDTISSSETVYMNVMTRLASYLKKLPFEFMQSGNNEALDLTELISNMPSLQ 318

Qy         68 -------FLTKGPSKLNDRADSRRSLWDQGNFPLI-IKNLK-----------IEDSDT-- 106
                   | |: | |  |       |  | | | | ::| |           | :: |
Db        319 LTPFSPVFGTEQPPKYEDVVQLSAS--DSGRFQTIELQNDKPNSRKMDTVQSIPNNSTNS 376

Qy        107 -YICEVEDQK--EEVQLLVFGLTAN-----SDTHLLQGQSLTLTLESPPGSSPSVQCRSP 158
                | || | :  : ||:    || |     |  |::    || :|       ::
Db        377 LYNLEVNDPRTLKAVQVQSQSLTMNPLENVSSDDLME----TLYIEEESDGKKALD---- 428

Qy        159 RGKNIQGG---KTLSVSQ--LELQDSGTW--TC-TVLQNQKKVEFKIDIVVLAFQKASSI 210
              :|:  : |   : | |::   |  :  |    | | :||:    |:   | |  :   |
Db        429 KGQKTENGPSHELLKVNEHRAEFPEHATHLKKCPTPMQNEIGKIFEKSFVNLPKEDCKSK 488

Qy        211 VYK-KEGEQVEFSFPLA-FTVEKLTGSGELWWQ------AERASSSKSWITFDLKNKEVS 262
              | | :||:| :|:   :   ::||    | : |       |  :  |:    :  ::
Db        489 VSKFEEGDQRDFTNSSSQEEIDKLLMDLESFSQKMETSLREPLAKGKNSNFLNSHSQLTG 548

Qy        263 VKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQL 322
                |  :|| ::    |:    |           ||  :|    |: :|   :::|  :
Db        549 QTLVDLEPKSKVSS--PIEKVSPSC---------LTRIIETNGHKIEEEDRALLLRILES 597

Qy        323 QKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVL------NPEAGMWQC------ 370
              ::   |:           :: ::    :|: :: : :|      : :| :  |
Db        598 IEDFAQEL----------VECKSSRGSLSQEKEMMQILQETLTTSSQANLSVCRSPVGD 646

Qy        371 -LLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALP 429
               :  ||::   :|:   : |     : || | |  : | |   | :|       |
Db        647 KAKDTTSAVLIQQTPEVIKIQNKPEKKPGTPLPPPATSPSSPRP--LSPVPHVNNVVNAP 704

Qy        430 AALAVISFLLGLGLGVAC 447
              :: :  |    ||   |
Db        705 LSINIPRFYFPEGLPDTC 722


RESULT 82
I12B_BOVIN
ID   I12B_BOVIN     STANDARD;      PRT;   327 AA.
AC   P46282;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Interleukin-12 beta chain precursor (IL-12B) (Cytotoxic lymphocyte
DE   maturation factor 40 kDa subunit) (CLMF p40).
GN   IL12B.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Holstein; TISSUE=Lymphoid;
RX   MEDLINE=95244628; PubMed=7727547;
RA   Zarlenga D.S., Canals A., Aschenbrenner R.A., Gasbarre L.C.;
RT   "Enzymatic amplification and molecular cloning of cDNA encoding the
RT   small and large subunits of bovine interleukin 12.";
RL   Biochim. Biophys. Acta 1270:215-217(1995).
CC   -!- FUNCTION: Cytokine that can act as a growth factor for activated T
CC       and NK cells, enhance the lytic activity of NK/lymphokine-
CC       activated Killer cells, and stimulate the production of IFN-gamma
CC       by resting PBMC (By similarity).
CC   -!- SUBUNIT: Disulfide-bonded heterodimer of 40 kDa and 35 kDa
CC       subunits.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Belongs to the type I cytokine family of receptors.
CC       Subfamily 3.
CC   -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U11815; AAA85792.1; -.
DR   GO; GO:0005615; C:extracellular space; NAS.
DR   GO; GO:0005143; F:interleukin-12 receptor binding; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0030546; F:receptor activator activity; ISS.
DR   GO; GO:0042095; P:interferon-gamma biosynthesis; IDA.
DR   GO; GO:0030101; P:natural killer cell activation; ISS.
DR   GO; GO:0042104; P:positive regulation of activated T-cell pro. . .; IDA.
DR   GO; GO:0042035; P:regulation of cytokine biosynthesis; ISS.
DR   GO; GO:0042093; P:T-helper cell differentiation; ISS.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003530; Hemtopoptn_L_F3.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 1.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS01354; HEMATOPO_REC_L_F3; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Cytokine; Glycoprotein; Immunoglobulin domain; Signal.
FT   SIGNAL        1     22       BY SIMILARITY.
FT   CHAIN        23    327       INTERLEUKIN-12 BETA CHAIN.
FT   DOMAIN       23    106       IG-LIKE C2-TYPE.
FT   DOMAIN      236    319       FIBRONECTIN TYPE-III.
FT   DISULFID     50     90       POTENTIAL.
FT   DISULFID    200    200       INTERCHAIN (PROBABLE).
FT   CARBOHYD    223    223       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   327 AA;  37050 MW;  D484DA36EF6F3D9F CRC64;

  Query Match             4.5%;  Score 105;  DB 1;  Length 327;
  Best Local Similarity   25.0%;  Pred. No. 3.7;
  Matches   69;  Conservative   38;  Mismatches   99;  Indels   70;  Gaps   17;

Qy        136 GQSLTLTLESPPGSSPSVQCRSPRGKNIQG-GKTLSVSQLELQDSGTWTCTVLQNQKKVE 194
              |::: || ::|      :   | :   : | ||||::   |  |:| :||      |  |
Db         43 GETVVLTCDTP--EEDGITWTSDQSSEVLGSGKTLTIQVKEFGDAGQYTC-----HKGGE 95

Qy        195 FKIDIVVLAFQKASSI----VYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKS 250
                ::|  :|   |    : | : |    || |    :  :|    ||
Db         96 ALSRSLLLLHKKEDGIWSTDILKDQKEPKAKSF-LKCEAKDYSGHFTCWWLTA------- 147

Qy        251 WITFDLKNKEVSVK--RVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKL 308
              |: |||    |||  | : ||:        :|   ||          |: | |
Db        148 -ISTDLK---FSVKSSRGSSDPR---------GVTCGAAL----------LSAE-KVSLE 183

Qy        309 HQEVNLVVMRATQLQKNLTCEVWGPTSPKLML---------SLKLENKEAKVSKRE--KP 357
              |:| |   :   : |:   |    | : : :|          || ||  :    |:  ||
Db        184 HREYNKYTV---ECQEGSAC----PAAEESLLIEVVVEAVHKLKYENYTSSFFIRDIIKP 236

Qy        358 VWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTP 393
                :|   :    | :| ||  | : :   |||||
Db        237 ----DPPKNLQLRPLKNSRQV--EVSWEYPDTWSTP 266


RESULT 83
OPCM_HUMAN
ID   OPCM_HUMAN     STANDARD;      PRT;   345 AA.
AC   Q14982;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Opioid binding protein/cell adhesion molecule precursor (OBCAM)
DE   (Opioid-binding cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Occipital cortex;
RX   MEDLINE=95237612; PubMed=7721093;
RA   Shark K.B., Lee N.M.;
RT   "Cloning, sequencing and localization to chromosome 11 of a cDNA
RT   encoding a human opioid-binding cell adhesion molecule (OBCAM).";
RL   Gene 155:213-217(1995).
CC   -!- FUNCTION: Binds opioids in the presence of acidic lipids; probably
CC       involved in cell contact.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC       family.
CC   -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L34774; AAA36387.1; -.
DR   PIR; JC4025; JC4025.
DR   Genew; HGNC:8143; OPCML.
DR   MIM; 600632; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004985; F:opioid receptor activity; TAS.
DR   GO; GO:0007155; P:cell adhesion; TAS.
DR   GO; GO:0008038; P:neuronal cell recognition; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal; Lipoprotein.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    322       OPIOID BINDING PROTEIN/CELL ADHESION
FT                                 MOLECULE.
FT   PROPEP      323    345       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       39    126       IG-LIKE C2-TYPE 1.
FT   DOMAIN      136    219       IG-LIKE C2-TYPE 2.
FT   DOMAIN      223    310       IG-LIKE C2-TYPE 3.
FT   DISULFID     57    115       POTENTIAL.
FT   DISULFID    157    202       POTENTIAL.
FT   DISULFID    244    296       POTENTIAL.
FT   CARBOHYD     44     44       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     70     70       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       322    322       GPI-anchor amidated asparagine
FT                                 (Potential).
SQ   SEQUENCE   345 AA;  38007 MW;  E7AD17BEA1AA3FF4 CRC64;

  Query Match             4.5%;  Score 105;  DB 1;  Length 345;
  Best Local Similarity   21.3%;  Pred. No. 3.9;
  Matches   48;  Conservative   40;  Mismatches   83;  Indels   54;  Gaps    8;

Qy         11 LLVLQLALLPAATQGNKVVLG-------------KKGDTVELTCTASQKKSIQFHWKNSN 57
              |:|: | ||      |  |  |               ::|::  | ||    : : :  | | :
Db         14 LVVVSLRLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVT-RVAWLNRS 72

Qy         58 QIKILGN-------QGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICE 110
                |   ||       :   |   |::                 : ::|:|: : |   | |
Db         73 TILYAGNDKWSIDPRVIILVNTPTQ---------------YSIMIQNVDVYDEGPYTCS 116
```

```
Qy        111 VE----DQKEEVQLLVFG----LTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKN 162
Db        117 VQTDNHPKTSRVHLIVQVPPQIMNISSDITVNEGSSVTLLCLAIGRPEPTVTWRHLSVKE 176

Qy        163 IQG----GKTLSVSQLELQDSGTWTCTVLQ-----NQKKVEFKID 198
Db        177 GQGFVSEDEYLEISDIKRDQSGEYECSALNDVAAPDVRKVKITVN 221


RESULT 84
RPOB_RICMA
ID   RPOB_RICMA     STANDARD;      PRT;  1373 AA.
AC   Q9RH43;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase
DE   beta chain) (RNA polymerase beta subunit).
GN   RPOB.
OS   Rickettsia massiliae.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=35791;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Mtu1;
RX   MEDLINE=99437773; PubMed=10508014;
RA   Drancourt M., Raoult D.;
RT   "Characterization of mutations in the rpoB gene in naturally
RT   rifampin-resistant Rickettsia species.";
RL   Antimicrob. Agents Chemother. 43:2400-2403(1999).
CC   -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC       of DNA into RNA using the four ribonucleoside triphosphates as
CC       substrates.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SUBUNIT: The enzyme consists of the sigma chain and the core
CC       enzyme which is composed of 2 alpha chains, 1 beta chain, and 1
CC       beta' chain.
CC   -!- SIMILARITY: Belongs to the RNA polymerase beta chain family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF076433; AAF22431.1; -.
DR   HSSP; Q9KWU7; 1HQM.
DR   InterPro; IPR007121; RNA_pol_B.
DR   InterPro; IPR007644; RNA_pol_Rpb2_1.
DR   InterPro; IPR007642; RNA_pol_Rpb2_2.
DR   InterPro; IPR007645; RNA_pol_Rpb2_3.
DR   InterPro; IPR007120; RNA_pol_Rpb2_6.
DR   InterPro; IPR007641; RNA_pol_Rpb2_7.
DR   Pfam; PF04563; RNA_pol_Rpb2_1; 1.
DR   Pfam; PF04561; RNA_pol_Rpb2_2; 1.
DR   Pfam; PF04565; RNA_pol_Rpb2_3; 1.
DR   Pfam; PF00562; RNA_pol_Rpb2_6; 1.
DR   Pfam; PF04560; RNA_pol_Rpb2_7; 1.
DR   PROSITE; PS01166; RNA_POL_BETA; 1.
KW   Transferase; DNA-directed RNA polymerase; Transcription.
SQ   SEQUENCE   1373 AA;  154278 MW;  C530FE629F99CACE CRC64;

  Query Match             4.5%;  Score 105;  DB 1;  Length 1373;
  Best Local Similarity   20.7%;  Pred. No. 24;
  Matches   67;  Conservative   50;  Mismatches  115;  Indels   92;  Gaps   13;

Qy         83 RRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKE-EVQLLVFGLTANSDTHLLQGQSLTL 141
Db        114 RLSIWD-------------IDEDTGTREIKGIKEQEVYMGDIPLMTKNGTFIINGTERVV 160

Qy        142 TLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKID--- 198
Db        161 V--SQMHRSPGVFFYHDEGKVHSSGKLLYSARV-IPYRGSWLDLEFDAKDVIYFRIDRKR 217

Qy        199 ----------------IVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQ 241
Db        218 KLYITTLLRAIGMSTEEIIKFYYNSVTYKLVKNKGWAVKF-IPQHITAHRLTS------- 269

Qy        242 AERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLAL 301
Db        270 ------------DLVDAD------TGNILLKAGQKI-----TPRLAKKYFGEG------ 299

Qy        302 EAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKL------MLSLKLENKEAKVSKRE 355
Db        300 ------LN---NILVAHETLIGKYLSEDLRDPTSDEVLAKIGEMITADMLNVINALKIKN 350

Qy        356 KPVWVLNPEAGMW--QCLLSDSGQ 377
Db        351 VNVLVINPQSGPYIRNTLFADKNQ 374


RESULT 85
ENV_BLVAU
ID   ENV_BLVAU      STANDARD;      PRT;   515 AA.
AC   P25057;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   ENV polyprotein precursor [Contains: Coat glycoprotein GP51; Coat
DE   protein GP30].
GN   ENV.
OS   Bovine leukemia virus (Australian isolate) (BLV).
OC   Viruses; Retroid viruses; Retroviridae; Deltaretrovirus.
OX   NCBI_TaxID=11903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90362060; PubMed=2167927;
RA   Coulston J., Naif H., Brandon R., Kumar S., Khan S., Daniel R.C.W.,
RA   Lavin M.F.;
RT   "Molecular cloning and sequencing of an Australian isolate of
RT   proviral bovine leukaemia virus DNA: comparison with other
RT   isolates.";
RL   J. Gen. Virol. 71:1737-1746(1990).
CC   -!- SIMILARITY: VERY STRONG, WITH BOVINE LEUKEMIA VIRUS ENV
CC       POLYPROTEIN FROM OTHER ISOLATES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D00647; BAA00545.1; -.
DR   PIR; JQ0556; VCLJGA.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; ENV_polyprotein; 1.
KW   Glycoprotein; Coat protein; Transmembrane; Polyprotein; Signal.
FT   SIGNAL        1     33
FT   CHAIN        34    301       COAT GLYCOPROTEIN GP51.
FT   CHAIN       302    515       COAT PROTEIN GP30.
FT   TRANSMEM    304    324       POTENTIAL.
FT   TRANSMEM    436    456       POTENTIAL.
FT   CARBOHYD     67     67       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    129    129       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    203    203       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    230    230       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    251    251       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    271    271       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    287    287       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    351    351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    398    398       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   515 AA;  58445 MW;  1CEB9D5A024A41E8 CRC64;

  Query Match             4.5%;  Score 104.5;  DB 1;  Length 515;
  Best Local Similarity   20.1%;  Pred. No. 7.2;
  Matches   68;  Conservative   48;  Mismatches  120;  Indels  103;  Gaps   17;

Qy        168 TLSVSQLEL-QDSGTWTCTV-LQNQ-------KKVEFKIDI-VVLAFQKASSIVYKKEGE 217
              :|::: | | |   || |:: | ||       :: :| | |  :|     |    |
Db         19 SLTLTLLALCQPIQTWRCSLSLGNQQWMTAYNQEAKFSISIDQILEAHNQSPFCAKSP-- 76

Qy        218 QVEFSFPLAFTVEKLTGSGELWW---QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQM 274
                  :|:: : |  :::|   | |   ::: :|:|              :|:
Db         77 --------RYTLDSVNGYPKIYWPPPQGRRRFGARAMVTYDC-----------EPRC-- 114

Qy        275 GKKLPLHLTLPQALPQYAGSGNLTL-----ALEAKTGKLHQEVNLVVMRATQLQK--NLT 327
                        | |:          | :|  |  :    :: :   |      ||
Db        115 ---------------PYVGADRFDCPHWDNASQADQGSFYVNHQILFLHLKQCHGIFTLT 159

Qy        328 CEVWGPTSPKLMLSL-KLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNI-- 384
              |:||   | :  || |: :       :    ||  |    |  ||: : :   :  |
Db        160 WEIWG-YDPLITFSLHKIPDPPQPDFPQLNSDWV--PSVRSWALLLNQTARAFPDCAICW 216

Qy        385 KVLPTWSTPVHPRASALPAPPTGSALPDPQ---------------------------TAS 417
              :  | |: :     : :   | |||| |                            :
Db        217 EPSPPWAPEILVYNKTISSSGPGLALPDAQIFWVNTSSFNTTQGWHHPSQRLLFNVSQGN 276

Qy        418 ALPDPP--------AASALPAAL---AVISFLLGLGLGV 445
              ||  ||        |:|| |  :    | :  ||| | |
Db        277 ALLLPPISLVNLSTASSAPPTRVRRSPVAALTLGLALSV 315


RESULT 86
ECTO_RAT
ID   ECTO_RAT       STANDARD;      PRT;   519 AA.
AC   P16573;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ecto-ATPase precursor (Cell-CAM 105) (C-CAM 105) (ATP-dependent
DE   taurocolate-carrier protein) (GP110).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF 50-68.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=89340561; PubMed=2527235;
RA   Lin S.-H., Guidotti G.;
RT   "Cloning and expression of a cDNA coding for a rat liver plasma
RT   membrane ecto-ATPase. The primary structure of the ecto-ATPase is
RT   similar to that of the human biliary glycoprotein I.";
RL   J. Biol. Chem. 264:14408-14414(1989).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=Sprague-Dawley, and Wistar; TISSUE=Liver;
RX   MEDLINE=93279310; PubMed=8504806;
RA   Edlund M., Gaardsvoll H., Bock E., Oebrink B.;
RT   "Different isoforms and stock-specific variants of the cell adhesion
RT   molecule C-CAM (cell-CAM 105) in rat liver.";
RL   Eur. J. Biochem. 213:1109-1116(1993).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=92344597; PubMed=1637321;
RA   Culic O., Huang Q., Flanagan D., Hixon D., Lin S.-H.;
RT   "Molecular cloning and expression of a new rat liver cell-CAM105
RT   isoform. Differential phosphorylation of isoforms.";
RL   Biochem. J. 285:47-53(1992).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Intestine;
RX   MEDLINE=94058980; PubMed=8240240;
RA   Cheung P.H., Culic O., Qiu Y., Earley K., Thompson N., Hixson D.C.,
RA   Lin S.-H.;
RT   "The cytoplasmic domain of C-CAM is required for C-CAM-mediated
RT   adhesion function: studies of a C-CAM transcript containing an
RT   unspliced intron.";
RL   Biochem. J. 295:427-435(1993).
RN   [5]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=90292222; PubMed=2141577;
RA   Aurivillius M., Hansen O.C., Lazrek M.B.S., Bock E., Oebrink B.;
RT   "The cell adhesion molecule Cell-CAM 105 is an ecto-ATPase and a
RT   member of the immunoglobulin superfamily.";
RL   FEBS Lett. 264:267-269(1990).
RN   [6]
RP   PARTIAL SEQUENCE.
RC   STRAIN=Wistar; TISSUE=Liver;
RX   MEDLINE=93292517; PubMed=8513803;
RA   Becker A., Lucka L., Kilian C., Kannicht C., Reutter W.;
RT   "Characterisation of the ATP-dependent taurocholate-carrier protein
RT   (gp110) of the hepatocyte canalicular membrane.";
RL   Eur. J. Biochem. 214:539-548(1993).
RN   [7]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=93123234; PubMed=8380406;
RA   Najjae S.M., Accili D., Philippe N., Jernberg J., Margolis R.,
RA   Taylor S.I.;
RT   "pp120/ecto-ATPase, an endogenous substrate of the insulin receptor
RT   tyrosine kinase, is expressed as two variably spliced isoforms.";
RL   J. Biol. Chem. 268:1201-1206(1993).
RN   [8]
RP   CHARACTERIZATION.
RX   MEDLINE=91354197; PubMed=1831973;
RA   Lin S.-H., Culic O., Flanagan D., Hixson D.C.;
RT   "Immunochemical characterization of two isoforms of rat liver ecto-
RT   ATPase that show an immunological and structural identity with a
RT   glycoprotein cell-adhesion molecule with Mr 105,000.";
RL   Biochem. J. 278:155-161(1991).
CC   -!- FUNCTION: THIS PROTEIN IS A CALCIUM-INDEPENDENT CELL ADHESION
CC       MOLECULE WITH HOMOPHILIC BINDING PROPERTIES. MAY PLAY A ROLE IN
CC       THE FORMATION AND MAINTENANCE OF THE SPECIALIZED MEMBRANE
CC       STRUCTURE OF THE APICAL SURFACE OF THE HEPATOCYTES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CANALICULAR DOMAIN
CC       OF HEPATOCYTE PLASMA MEMBRANES.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=P16573-1; Sequence=Displayed;
CC       Name=Short; Synonyms=C-CAM3;
CC         IsoId=P16573-2; Sequence=VSP_002504, VSP_002505;
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN EPITHELIA, VESSEL ENDOTHELIA,
CC       LEUKOCYTES AND PLATELETS.
CC   -!- PTM: PHOSPHORYLATED ON SERINE RESIDUES IN BOTH ISOFORMS. ON SERINE
CC       AND TYROSINE RESIDUES IN CYTOPLASMIC REGION OF THE LONG ISOFORM.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC   -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J04963; AAA41104.1; -.
DR   EMBL; X71122; CAA50435.1; -.
DR   EMBL; Z12019; CAA78054.1; -.
DR   EMBL; M92848; AAA16783.1; ALT_TERM.
DR   PIR; A44783; A44783.
```

```
DR   PIR; S23969; S23969.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Cell adhesion; Transmembrane; Glycoprotein; Immunoglobulin domain;
KW   Repeat; Signal; Phosphorylation; Alternative splicing;
KW   Pyrrolidone carboxylic acid.
FT   SIGNAL        1     34       POTENTIAL.
FT   CHAIN        35    519       ECTO-ATPASE.
FT   DOMAIN       35    422       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    423    448       POTENTIAL.
FT   DOMAIN      449    519       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       42    140       IG-LIKE V-TYPE.
FT   DOMAIN      147    232       IG-LIKE C2-TYPE 1.
FT   DOMAIN      237    317       IG-LIKE C2-TYPE 2.
FT   DOMAIN      325    403       IG-LIKE C2-TYPE 3.
FT   DISULFID    167    215       BY SIMILARITY.
FT   DISULFID    259    299       BY SIMILARITY.
FT   DISULFID    344    392       BY SIMILARITY.
FT   MOD_RES      35     35       PYRROLIDONE CARBOXYLIC ACID
FT                                (BY SIMILARITY).
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    113    113       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    172    172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    173    173       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    197    197       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    224    224       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    256    256       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    288    288       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    331    331       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    374    374       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    455    458       GSDH -> SGSF (in isoform Short).
FT                                /FTId=VSP_002504.
FT   VARSPLIC    459    519       Missing (in isoform Short).
FT                                /FTId=VSP_002505.
FT   VARIANT      49     49       S -> K.
FT   VARIANT      55     55       T -> A.
FT   VARIANT      70     70       V -> G.
FT   VARIANT      73     76       TGLN -> LNPD.
FT   VARIANT      86     86       S -> D.
FT   VARIANT      88     88       T -> M.
FT   VARIANT      90     90       Q -> K.
FT   VARIANT      92     92       E -> G.
FT   VARIANT      99     99       V -> E.
FT   VARIANT     118    119       GP -> RA.
FT   VARIANT     125    125       I -> F.
FT   VARIANT     127    127       K -> Q.
SQ   SEQUENCE   519 AA;  57263 MW;  04143928D7975776 CRC64;

  Query Match             4.5%;  Score 104.5;  DB 1;  Length 519;
  Best Local Similarity   19.0%;  Pred. No. 7.2;
  Matches   99;  Conservative   84;  Mismatches  177;  Indels  161;  Gaps   27;

Qy          3 RG-VPFRHLLLVLQL--------------ALLPAATQGNKVVLGKKGDTVELTCTASQK 46
              || :|:| |||   |               |: |   : : |:|        ||    |:
Db         10 RGQIPWRGLLLTASLLTYWSPLTTAQVTVDAVPPNVVEESSVLL--------LTHNLPQE 61

Qy         47 KSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRR-SLWDQGNFPLIIKNLKIEDSD 105
                : |:|       :      ::     :   : | | | ::: |:  |   |:|:   |
Db         62 FQV-FYWYKVTTTGLNSEIARYIRSSNTSQTEPAYSGRVTIYSNGS--LFFQNVNKTDEG 118

Qy        106 TYICEVEDQ-----KEEVQLLVF------GLTANSDTHLLQGQSLTLTLESPPGSSPSVQ 154
               |   | |:     :  ||  |:        :| | ::: ::|:   ::|    | :: :
Db        119 PYTLSVIDKQFNPIQTSVQFRVYPALQKPNVTGN-NSNPMEGEPF-VSLMCEPYTNNTSY 176

Qy        155 CRSPRGKNI---------QGGKTLSVSQLELQDSGTWTCTVLQN---QKKVEFKIDIV-- 200
                |  |:::          :| :||::  :   | | : |          :   | :|::
Db        177 LWSRNGESLSEGDRVTFSEGNRTLTLLNVRRTDKGYYECEARNPATFNRSDPFNLDVIYG 236

Qy        201 ----VLAFQKASSIVYKKEGEQVEFSF-----PLA----FTVEKL-TGSGELWWQAERAS 246
                  |:    :   :|  :|  :  |       | |        |||  | | ||:      :
Db        237 PDAPVI----SPPDIYLHQGSNLNLSCHADSNPPAQYFWLINEKLQTSSQELFISNITTN 292

Qy        247 SSKSWITFDLKN----------KEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGN 296
                :| ::  | : |          | ::|      | :|:       |             |:
Db        293 NSGTYACF-VNNTVTGLSRTTVKNITVFEPVTQPSIQITNTTVKEL-----------GS 339

Qy        297 LTLALEAK-TGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLML----SLKLENKEAKV 351
                :||   :| ||        | :|           |   |  | |    :|  :|   ::
Db        340 VTLTCFSKDTG--------VSVR-----------WLFNSQSLQLTDRMTLSQDNSTLRI 379

Qy        352 S--KREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSA 409
                 |||        :|| :|| :|:                                  |
Db        380 DPIKRE--------DAGDYQCEISN----------------------------PVSFR 401

Qy        410 LPDPQTASALPDPPAASALPAALAVISFLLGLGLGVACVLA 450
                :  |      :|||   ::  :  |:   ::|   ||| : |
Db        402 ISHPIKLDVIPDPTQGNSGLSEGAIAGIVIGSVAGVALIAA 442


RESULT 87
PTK7_HUMAN
ID   PTK7_HUMAN     STANDARD;      PRT;  1070 AA.
AC   Q13308; Q13417;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tyrosine-protein kinase-like 7 precursor (Colon carcinoma kinase-4)
DE   (CCK-4).
GN   PTK7 OR CCK4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon carcinoma, and Placenta;
RX   MEDLINE=96074849; PubMed=7478540;
RA   Mossie K., Jallal B., Alves F., Sures I., Plowman G.D., Ullrich A.;
RT   "Colon carcinoma kinase-4 defines a new subclass of the receptor
RT   tyrosine kinase family.";
RL   Oncogene 11:2179-2184(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fibroblast;
RX   MEDLINE=97037064; PubMed=8882711;
RA   Park S.-K., Lee H.-S., Lee S.-T.;
RT   "Characterization of the human full-length PTK7 cDNA encoding a
RT   receptor protein tyrosine kinase-like molecule closely related to
RT   chick KLG.";
RL   J. Biochem. 119:235-239(1996).
RN   [3]
RP   SEQUENCE FROM N.A., AND REVISION TO 834.
RA   Lee S.-T., Park S.-K., Lee H.-S., Ji A.R., Jung J.W.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   CARBOHYDRATE-LINKAGE SITE ASN-646.
RX   MEDLINE=22660472; PubMed=12754519;
RA   Zhang H., Li X.-J., Martin D.B., Aebersold R.;
RT   "Identification and quantification of N-linked glycoproteins using
RT   hydrazide chemistry, stable isotope labeling and mass spectrometry.";
RL   Nat. Biotechnol. 21:660-666(2003).
CC   -!- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY
CC       THE CATALYTIC ACTIVITY OF TYROSINE KINASE. MAY BE CONNECTED TO THE
CC       PATHOPHYSIOLOGY OF COLON CARCINOMAS AND/OR MAY REPRESENT A TUMOR
CC       PROGRESSION MARKER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, LIVER, PANCREAS,
```

```
CC       KIDNEY, PLACENTA AND MELANOCYTES. WEAKLY EXPRESSED IN THYROID
CC       GLAND, OVARY, BRAIN, HEART AND SKELETAL MUSCLE. ALSO EXPRESSED IN
CC       ERYTHROLEUKEMIA CELLS. BUT NOT EXPRESSED IN COLON.
CC   -!- SIMILARITY: Belongs to the Tyr family of protein kinases. Insulin
CC       receptor subfamily.
CC   -!- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U33635; AAA87565.1; -.
DR   EMBL; U40271; AAC50484.2; -.
DR   EMBL; AF447176; AAL39062.1; -.
DR   EMBL; AF447157; AAL39062.1; JOINED.
DR   EMBL; AF447158; AAL39062.1; JOINED.
DR   EMBL; AF447162; AAL39062.1; JOINED.
DR   EMBL; AF447164; AAL39062.1; JOINED.
DR   EMBL; AF447167; AAL39062.1; JOINED.
DR   EMBL; AF447170; AAL39062.1; JOINED.
DR   EMBL; AF447171; AAL39062.1; JOINED.
DR   EMBL; AF447173; AAL39062.1; JOINED.
DR   EMBL; AF447174; AAL39062.1; JOINED.
DR   EMBL; AF447175; AAL39062.1; JOINED.
DR   HSSP; P08631; 1AD5.
DR   Genew; HGNC:9618; PTK7.
DR   MIM; 601890; -.
DR   GO; GO:0005886; C:plasma membrane; TAS.
DR   GO; GO:0005888; C:proteoglycan integral to plasma membrane; TAS.
DR   GO; GO:0004714; F:transmembrane receptor protein tyrosine kin. . .; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002011; RecepttyrkinsII.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   Pfam; PF00047; ig; 7.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00408; IGc2; 5.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 7.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; FALSE_NEG.
KW   Receptor; Transmembrane; Signal; Glycoprotein; Cell adhesion;
KW   Immunoglobulin domain; Repeat.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1070       TYROSINE-PROTEIN KINASE-LIKE 7.
FT   DOMAIN       31    704       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    705    725       POTENTIAL.
FT   DOMAIN      726   1070       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31    120       IG-LIKE C2-TYPE 1.
FT   DOMAIN      128    218       IG-LIKE C2-TYPE 2.
FT   DOMAIN      225    317       IG-LIKE C2-TYPE 3.
FT   DOMAIN      309    407       IG-LIKE C2-TYPE 4.
FT   DOMAIN      412    497       IG-LIKE C2-TYPE 5.
FT   DOMAIN      503    586       IG-LIKE C2-TYPE 6.
FT   DOMAIN      578    680       IG-LIKE C2-TYPE 7.
FT   DOMAIN      796   1066       PROTEIN KINASE; INACTIVE.
FT   DISULFID     53    101       BY SIMILARITY.
FT   DISULFID    150    200       BY SIMILARITY.
FT   DISULFID    246    301       BY SIMILARITY.
FT   DISULFID    343    391       BY SIMILARITY.
FT   DISULFID    433    481       BY SIMILARITY.
FT   DISULFID    524    570       BY SIMILARITY.
FT   DISULFID    613    664       BY SIMILARITY.
FT   CARBOHYD    116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    175    175       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    184    184       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    214    214       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    268    268       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    283    283       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    405    405       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    463    463       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    567    567       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    646    646       N-LINKED (GLCNAC. . .).
FT   CONFLICT     92     92       P -> R (IN REF. 2 AND 3).
FT   CONFLICT    147    147       K -> T (IN REF. 2 AND 3).
FT   CONFLICT    207    207       S -> G (IN REF. 2 AND 3).
FT   CONFLICT    495    496       VL -> RV (IN REF. 2 AND 3).
FT   CONFLICT    515    515       G -> E (IN REF. 2 AND 3).
FT   CONFLICT    881    881       E -> G (IN REF. 2 AND 3).
FT   CONFLICT    969    969       A -> P (IN REF. 2 AND 3).
FT   CONFLICT    992    992       S -> F (IN REF. 2 AND 3).
SQ   SEQUENCE   1070 AA;  118260 MW;  47CDF25B8E3698A5 CRC64;

  Query Match             4.5%;  Score 104.5;  DB 1;  Length 1070;
  Best Local Similarity   20.8%;  Pred. No. 18;
  Matches   86;  Conservative   58;  Mismatches  147;  Indels  123;  Gaps   22;

Qy         89 QGNFPLIIKNLKIEDSDTYICEVED----QKEEVQLLV----FGLTANSDTHLLQGQSLT 140
              |    |:: |:   |:  | |   :    ::::| : |      |    |: | :|:
Db        371 QKGHELVLANIAESDAGVYTCHAANLAGQRRQDVNITVATVPSWLKKPQDSQLEEGKPGY 430

Qy        141 LTLESPPGSSPSVQ--------CRSPRGKNIQGGKTLSVSQLELQDSGTW-TC------- 184
              |   :     |:|             | :  : | || :: :|: | |||  |
Db        431 LDCLTQATPKPTVVWYRNQMLISEDSRFEVFKNG-TLRINSVEVYD-GTWYRCMSSTPAG 488

Qy        185 -----TVLQNQKKVEF---KIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSG 236
                  |||  :|::|         : | | :::     |             || |
Db        489 SIEAQAVLQVLEKLKFTPPPQPQQCMGFDKEATVPCSATGR-----------EKPT--- 533

Qy        237 ELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGN 296
               : |:   ||   |:| :    :   |||:|                        :||
Db        534 -IKWERADGSSLPEWVTDNAGT--LHFARVTRD----------------------DAGN 567

Qy        297 LT-LALEAKTGKLHQEVNLVVM----------RATQLQKN---LTCEVWGPTSPKLMLSL 342
              | :|     |::  | | |           | |  | :   | ||  |   || ::
Db        568 YTCIASNGPQGQIRAHVQLTVAVFITFKVEPERTTVYQGHTALLQCEAQG--DPKPLIQW 625

Qy        343 KLENKEAKVSKREKPVWVLN----------PE-AGMWQCLLSDSGQVLLESNIKVLPTWS 391
              | :::    :|    : :            || :| : |:  :|       |||
Db        626 KGKDRILDPTKLGPRMHIFQNGSLVIHDVAPEDSGRYTCIAGNS------CNIK---HTE 676

Qy        392 TPVHPRASALPAPPTGSALPDP----QTASALPDPPAASALPAALAVISFLLGL 441
              |::     :|   | | |     ||          :: ||:| |  :|||
Db        677 APLYVVDKPVPEESEGPGSPPPYKMIQTIGL--------SVGAAVAYIIAVLGL 722


RESULT 88
E2F2_HUMAN
ID   E2F2_HUMAN     STANDARD;      PRT;   437 AA.
AC   Q14209;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Transcription factor E2F2 (E2F-2).
GN   E2F2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94067141; PubMed=8246995;
RA   Ivey-Hoyle M., Conroy R., Huber H.E., Goodhart P.J., Oliff A.,
RA   Heimbrook D.C.;
RT   "Cloning and characterization of E2F-2, a novel protein with the
```

```
RT   biochemical properties of transcription factor E2F.";
RL   Mol. Cell. Biol. 13:7802-7812(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Livingston R.J., Braun A.C., Montoya M.A., Chung M.-W.,
RA   Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA   Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Pearce A.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Transcription activator that binds DNA cooperatively
CC       with DP proteins through the E2 recognition site, TTTCC/GCGC,
CC       found in the promoter region of a number of genes whose products
CC       are involved in cell cycle regulation or in DNA replication. The
CC       DRTF1/E2F complex functions in the control of cell-cycle
CC       progression from g1 to s phase. E2F-2 binds specifically to RB1
CC       protein, in a cell-cycle dependent manner.
CC   -!- SUBUNIT: Component of the DRTF1/E2F transcription factor complex.
CC       Forms heterodimers with DP family members. The E2F-2 complex binds
CC       specifically hypophosphorylated retinoblastoma protein RB1. During
CC       the cell cycle, RB1 becomes phosphorylated in mid-to-late G1
CC       phase, detaches from the DRTF1/E2F complex, rendering E2F
CC       transcriptionally active. Viral oncoproteins, notably E1A, T-
CC       antigen and HPV E7, are capable of sequestering RB protein, thus
CC       releasing the active complex.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Highest level of expression is found in
CC       placenta, low levels are found in lung. Found as well in many
CC       immortalized cell lines derived from tumor samples.
CC   -!- PTM: Phosphorylated by CDK2 and cyclin A-CDK2 in the S-phase (By
CC       similarity).
CC   -!- SIMILARITY: Belongs to the E2F/DP family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L22846; AAA16890.1; -.
DR   EMBL; AF518877; AAM54044.1; -.
DR   EMBL; AL021154; CAA15949.1; -.
DR   PIR; A54595; A54595.
DR   PDB; 1N4M; 07-JAN-03.
DR   TRANSFAC; T01544; -.
DR   Genew; HGNC:3114; E2F2.
DR   GK; Q14209; -.
DR   MIM; 600426; -.
DR   GO; GO:0003677; F:DNA binding; TAS.
DR   GO; GO:0005515; F:protein binding; TAS.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
DR   GO; GO:0006367; P:transcription initiation from Pol II promoter; TAS.
DR   InterPro; IPR003316; E2F_TDP.
DR   Pfam; PF02319; E2F_TDP; 1.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW   Phosphorylation; Cell cycle; 3D-structure.
FT   DOMAIN       65    105       CYCLIN A/CDK2 BINDING (POTENTIAL).
FT   DNA_BIND     107   196       POTENTIAL.
FT   DOMAIN       155   176       LEUCINE-ZIPPER.
FT   DOMAIN       160   196       DEF BOX.
FT   DOMAIN       197   289       DIMERIZATION (POTENTIAL).
FT   DOMAIN       359   437       TRANSACTIVATION (POTENTIAL).
FT   DOMAIN       360   363       POLY-PRO.
FT   DOMAIN       410   427       RETINOBLASTOMA PROTEIN ASSOCIATION
FT                                (POTENTIAL).
SQ   SEQUENCE   437 AA;  47505 MW;  60541F4235507005 CRC64;

  Query Match             4.5%;  Score 104;  DB 1;  Length 437;
  Best Local Similarity   22.3%;  Pred. No. 6.3;

  Matches   75;  Conservative   48;  Mismatches  145;  Indels   68;  Gaps   15;

Qy        145 SPPGSSPSVQC-----RSPRGKNIQ---GGKTLSVSQLELQDSGTWTCTVLQNQKKVEFK 196
              :|| ::|   |       | |: ::    |:  :  :|:|:  |          |    :
Db         53 APPAAAPGT-CLDATPHGPEGQVVRCLPAGRLPAKRKLDLEGIGRPVVPEFPTPKGKCIR 111

Qy        197 IDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKL--------TGSGELWWQAERASSS 248
              :| :       |     |  ||:  :   |    :|          |  :| | ||
Db        112 VDGL------PSPKTPKSPGEKTRYDTSLGLLTKKFIYLLSESEDGVLDLNWAAEVLDVQ 165

Qy        249 KSWITFDLKNKEVSVKRVT---------------QDP-----KLQMGKKLPLHLTLPQAL 288
              |  | :|: |    :: :                 :||     : |:|::|   :   |||
Db        166 KRRI-YDITNVLEGIQLIRKKAKNNIQWVGRGMFEDPTRPGKQQQLGQELKELMNTEQAL 224

Qy        289 PQYAGSGNLTL--ALEAKTGKLHQEVNLVVMRAT-QLQKNLTCEVWGPTSPKL------- 338
              |   | :|:     | |  |    |    :||    ::     |  |   :|
Db        225 DQLIQSCSLSFKHLTEDKANKRLAYVTYQDIRAVGNFKEQTVIAVKAPPQTRLEVPDRTE 284

Qy        339 -MLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTP---V 394
               | : |:: :  :      |  |  |::   : | | |   |  |     |: ||   :
Db        285 DNLQIYLKSTQGPIEVYLCPEEVQEPDSPSEEPLPSTS--TLCPSPDSAQPSSSTDPSIM 342

Qy        395 HPRASALPAPPTGSALPDPQTASALPDPPAASALPA 430
              | ||::|||      | || |   | ||:   | |
Db        343 EPTASSVPAP-----APTPQQA---PPPPSLVPLEA 370


RESULT 89
BUTY_MOUSE
ID   BUTY_MOUSE     STANDARD;      PRT;   524 AA.
AC   Q62556; P97392;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Butyrophilin precursor (BT) (Butyrophilin subfamily 1 member A1).
GN   BTN1A1 OR BTN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129; TISSUE=Mammary gland;
RX   MEDLINE=97148936; PubMed=8995761;
RA   Ogg S.L., Komaragiri M.V.S., Mather I.H.;
RT   "Structural organization and mammary-specific expression of the
RT   butyrophilin gene.";
RL   Mamm. Genome 7:900-905(1996).
RN   [2]
RP   SEQUENCE OF 39-487 FROM N.A.
RC   TISSUE=Mammary gland;
RX   MEDLINE=96125722; PubMed=8541302;
RA   Ishii T., Aoki N., Noda A., Adachi T., Nakamura R., Matsuda T.;
RT   "Carboxy-terminal cytoplasmic domain of mouse butyrophilin
RT   specifically associates with a 150-kDa protein of mammary epithelial
RT   cells and milk fat globule membrane.";
RL   Biochim. Biophys. Acta 1245:285-292(1995).
CC   -!- FUNCTION: May function in the secretion of milk-fat droplets. It
CC       may act as a specific membrane-associated receptor for the
CC       association of cytoplasmic droplets with the apical plasma
CC       membrane (By similarity).
CC   -!- SUBUNIT: Seems to associate with xanthine dehydrogenase/oxidase.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: Expressed in mammary tissue and secreted in
CC       association with the milk-fat-globule membrane during lactation.
CC   -!- DEVELOPMENTAL STAGE: Expression increases during the last half of
CC       pregnancy and is maximal during lactation.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. BTN/MOG
CC       family.
CC   -!- SIMILARITY: Contains 2 immunoglobulin-like V-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U67065; AAB51034.1; -.
DR   EMBL; S80642; AAB35893.1; -.
DR   MGD; MGI:103118; Btn1a1.
DR   InterPro; IPR001870; B302.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR006574; PRY.
DR   InterPro; IPR003877; SPRY_receptor.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00622; SPRY; 1.
DR   SMART; SM00406; IGv; 1.
DR   SMART; SM00589; PRY; 1.
DR   SMART; SM00449; SPRY; 1.
DR   PROSITE; PS50835; IG_LIKE; 2.
KW   Transmembrane; Glycoprotein; Immunoglobulin domain; Signal; Repeat.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    524       BUTYROPHILIN.
FT   DOMAIN       27    247       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    248    268       POTENTIAL.
FT   DOMAIN      269    524       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       29    141       IG-LIKE V-TYPE 1.
FT   DOMAIN      149    235       IG-LIKE V-TYPE 2.
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    216    216       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     46     46       D -> DD (IN REF. 2).
FT   CONFLICT    117    117       V -> F (IN REF. 2).
FT   CONFLICT    191    191       E -> D (IN REF. 2).
FT   CONFLICT    210    210       R -> S (IN REF. 2).
FT   CONFLICT    363    363       V -> E (IN REF. 2).
FT   CONFLICT    408    408       T -> K (IN REF. 2).
FT   CONFLICT    413    414       SL -> FF (IN REF. 2).
FT   CONFLICT    420    423       PRRV -> LAEY (IN REF. 2).
FT   CONFLICT    492    509       DIPLSPLGEGCTSGDKDT -> GHSLVPAGGRLYFWRQRH
FT                                (IN REF. 2).
SQ   SEQUENCE   524 AA;  58406 MW;  333F4DE2C7704480 CRC64;

  Query Match             4.5%;  Score 104;  DB 1;  Length 524;
  Best Local Similarity   22.5%;  Pred. No. 7.9;
  Matches   73;  Conservative   52;  Mismatches  108;  Indels   92;  Gaps   17;

Qy         11 LLVLQLALLPAA-----TQGNKVVLGKKGDTVELTCTASQKKSIQF----HWKNSNQIKI 61
              | ||||  | :|     |   : ||   |   ||||  |   | ::    :: :   :
Db         16 LTVLQLPTLDSAAPFDVTAPQEPVLALVGSDAELTCGFSPNASSEYMELLWFRQTRSTAV 75

Qy         62 L------GNQGSFLT--KGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED 113
              |      :|  :|  :| : |     :   | | |   |:|::::: |   | |  :|
Db         76 LLYRDGQEQEGQQMTEYRGRATL-----ATAGLLD-GRATLLIRDVRVSDQGEYRCLFKD 129

Qy        114 ----QKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGS---SPSVQCRS--------- 157
                  ::  | | |  : ::    :   ::  : ||         | || |:
Db        130 NDDFEEAAVYLKVAAVGSDPQISMTVQENGEMELECTSSGWYPEPQVQWRTGNREMLPST 189

Qy        158 --PRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKE 215
               :  | :|  |::|| : ::||          | :   |  ::|          :   .
Db        190 SESKKHNEEGLFTVAVSMM-IRDSSI---------KNMSCCIQNILLG----------Q 228

Qy        216 GEQVEFSFPLAFTVEKLT----------------GSGELWWQAERASSSKSWITFDLK 257
              |::|| | |  | | :||                ||    |:  :  ||       |:
Db        229 GKEVEISLPAPF-VPRLTPWIVAVAIILLALGFLTIGSIFFTWKLYKERSS-------LR 280

Qy        258 NKEV-SVKRVTQDPKLQMGKKLPLH 281
               ||  | :|: :: :    ||  ||
Db        281 KKEFGSKERLLEELRC---KKTVLH 302


RESULT 90
CD22_MOUSE
ID   CD22_MOUSE     STANDARD;      PRT;   862 AA.
AC   P35329; Q9JHL2; Q9JJX9; Q9JJY0; Q9JJY1; Q9R056; Q9R094; Q9WU51;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   B-cell receptor CD22 precursor (Leu-14) (B-lymphocyte cell adhesion
DE   molecule) (BL-CAM) (Siglec-2).
GN   CD22 OR LYB-8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBA/2J, and BALB/c; TISSUE=Liver;
RX   MEDLINE=93315834; PubMed=8100843;
RA   Law C.-L., Torres R.M., Sundberg H.A., Parkhouse R.M.,
RA   Brannan C.I., Copeland N.G., Jenkins N.A., Clark E.A.;
RT   "Organization of the murine Cd22 locus. Mapping to chromosome 7 and
RT   characterization of two alleles.";
RL   J. Immunol. 151:175-187(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6, BXSB, and MRL;
RX   MEDLINE=99432012; PubMed=10501843;
RA   Lajaunias F., Ibnou-Zekri N., Fossati Jimack L., Chicheportiche Y.,
RA   Parkhouse R.M., Mary C., Reininger L., Brighouse G., Izui S.;
RT   "Polymorphisms in the Cd22 gene of inbred mouse strains.";
RL   Immunogenetics 49:991-995(1999).
RN   [3]
RP   SEQUENCE OF 1-178 FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   STRAIN=NZW; TISSUE=Spleen;
RX   MEDLINE=20432275; PubMed=10975807;
RA   Mary C., Laporte C., Parzy D., Santiago M.L., Stefani F.,
RA   Lajaunias F., Parkhouse M.E., O'Keefe T.L., Neuberger M.S., Izui S.,
RA   Reininger L.;
RT   "Dysregulated expression of the Cd22 gene as a result of a short
RT   interspersed nucleotide element insertion in Cd22alpha lupus-prone
RT   mice.";
RL   J. Immunol. 165:2987-2996(2000).
RN   [4]
RP   SIALIC ACID BINDING.
RX   MEDLINE=95179521; PubMed=7533044;
RA   Kelm S., Pelz A., Schauer R., Filbin M.T., Tang S., de Bellard M.E.,
RA   Schnaar R.L., Mahoney J.A., Hartnell A., Bradfield P., Crocker P.R.;
RT   "Sialoadhesin, myelin-associated glycoprotein and CD22 define a new
RT   family of sialic acid-dependent adhesion molecules of the
RT   immunoglobulin superfamily.";
RL   Curr. Biol. 4:965-972(1994).
RN   [5]
RP   INTERACTION WITH GRB2; SYK; PIK3R1/PIK3R2 AND PLCG1, PHOSPHORYLATION
RP   OF TYR-777; TYR-822; TYR-837 AND TYR-857, AND MUTAGENESIS OF TYR-822.
RX   MEDLINE=99303650; PubMed=10373493;
RA   Yohannan J., Wienands J., Coggeshall K.M., Justement L.B.;
RT   "Analysis of tyrosine phosphorylation-dependent interactions between
RT   stimulatory effector proteins and the B cell co-receptor CD22.";
RL   J. Biol. Chem. 274:18769-18776(1999).
RN   [6]
RP   INTERACTION WITH GRB2; SHC1 AND INPP5D, AND PHOSPHORYLATION BY LYN.
RX   MEDLINE=20298808; PubMed=10748054;
RA   Poe J.C., Fujimoto M., Jansen P.J., Miller A.S., Tedder T.F.;
RT   "CD22 forms a quaternary complex with SHIP, Grb2, and Shc. A pathway
RT   for regulation of B lymphocyte antigen receptor-induced calcium
RT   flux.";
RL   J. Biol. Chem. 275:17420-17427(2000).
CC   -!- FUNCTION: Mediates B-cell B-cell interactions. May be involved in
CC       the localization of B-cells in lymphoid tissues. Binds sialylated
CC       glycoproteins; one of which is CD45. Preferentially binds to
CC       alpha2,6-linked sialic acid. The sialic acid recognition site can
CC       be masked by cis interactions with sialic acids on the same cell
CC       surface. Upon ligand induced tyrosine phosphorylation in the
CC       immune response seems to be involved in regulation of B cell
CC       antigen receptor signaling. Plays a role in positive regulation
```

```
CC       through interaction with Src family tyrosine kinases and may also
CC       act as an inhibitory receptor by recruiting cytoplasmic
CC       phosphatases via their SH2 domains that block signal transduction
CC       through dephosphorylation of signaling molecules.
CC   -!- SUBUNIT: Interacts with SYK, PIK3R1/PIK3R2, PLCG1, SHC1, INPP5D
CC       and GRB2 upon phosphorylation. May form a complex with
CC       INPP5D/SHIP, GRB2 and SHC1. Interacts with PTPN6/SHP-1 upon
CC       phosphorylation (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1; Synonyms=CD22-beta;
CC         IsoId=P35329-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P35329-2; Sequence=VSP_002532;
CC       Name=3;
CC         IsoId=P35329-3; Sequence=VSP_002533;
CC   -!- TISSUE SPECIFICITY: B lymphocytes.
CC   -!- DOMAIN: Contains 3 copies of a cytoplasmic motif that is referred
CC       to as the immunoreceptor tyrosine-based inhibitor motif (ITIM).
CC       This motif is involved in downmodulation of cellular responses.
CC       The phosphorylated ITIM motif binds to the SH2 domain of
CC       PTPN6/SHP-1.
CC   -!- PTM: Phosphorylated on tyrosine residues by LYN (Probable).
CC   -!- PTM: Phosphorylation on Tyr-822 is involved in the binding to
CC       GRB2.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC
CC       (sialic acid binding Ig-like lectin) family.
CC   -!- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L16928; AAA02562.1; -.
DR   EMBL; AF115401; AAD30392.1; -.
DR   EMBL; AF115400; AAD30391.1; -.
DR   EMBL; AF102134; AAF02417.1; -.
DR   EMBL; AJ250676; CAB85609.1; -.
DR   EMBL; AJ250677; CAB85610.1; -.
DR   EMBL; AJ250678; CAB85611.1; -.
DR   EMBL; AJ250679; CAB85612.1; -.
DR   EMBL; AJ250680; CAB85613.1; -.
DR   EMBL; AJ250682; CAB85615.1; -.
DR   EMBL; AJ250683; CAB85616.1; -.
DR   PIR; I49583; I49583.
DR   MGD; MGI:88322; Cd22.
DR   GO; GO:0005887; C:integral to plasma membrane; IPI.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00408; IGc2; 4.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   Cell adhesion; Lectin; Transmembrane; Signal; Glycoprotein;
KW   Phosphorylation; Immunoglobulin domain; Repeat; Alternative splicing.
FT   SIGNAL       1    21       POTENTIAL.
FT   CHAIN       22   862       B-CELL RECEPTOR CD22.
FT   DOMAIN      22   702       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   703   721       POTENTIAL.
FT   DOMAIN     722   862       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      22   142       IG-LIKE V-TYPE.
FT   DOMAIN     147   244       IG-LIKE C2-TYPE 1.
FT   DOMAIN     251   341       IG-LIKE C2-TYPE 2.
FT   DOMAIN     346   429       IG-LIKE C2-TYPE 3.
FT   DOMAIN     434   515       IG-LIKE C2-TYPE 4.
FT   DOMAIN     520   597       IG-LIKE C2-TYPE 5.
FT   DOMAIN     608   691       IG-LIKE C2-TYPE 6.
FT   SITE       775   780       ITIM MOTIF 1.
FT   SITE       835   840       ITIM MOTIF 2.
FT   SITE       855   860       ITIM MOTIF 3.
FT   DISULFID    41   171       BY SIMILARITY.
FT   DISULFID    46   106       BY SIMILARITY.
FT   DISULFID   165   229       BY SIMILARITY.
FT   DISULFID   272   324       BY SIMILARITY.
FT   DISULFID   368   411       BY SIMILARITY.
FT   DISULFID   457   499       BY SIMILARITY.
FT   DISULFID   544   586       BY SIMILARITY.
FT   DISULFID   631   674       BY SIMILARITY.
FT   MOD_RES    777   777       PHOSPHORYLATION
FT                              (INVOLVED IN BINDING TO SYK) (PROBABLE).
FT   MOD_RES    822   822       PHOSPHORYLATION (PROBABLE).
FT   MOD_RES    837   837       PHOSPHORYLATION
FT                              (INVOLVED IN BINDING TO SYK) (PROBABLE).
FT   MOD_RES    857   857       PHOSPHORYLATION
FT                              (INVOLVED IN BINDING TO SYK, PLCG2 AND
FT                              PIK3R1/PIK3R2) (PROBABLE).
FT   CARBOHYD   105   105       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   116   116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   139   139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   168   168       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   265   265       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   275   275       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   378   378       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   408   408       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   460   460       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   561   561       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   589   589       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    54   115       Missing (in isoform 2).
FT                              /FTId=VSP_002532.
FT   VARSPLIC   100   133       Missing (in isoform 3).
FT                              /FTId=VSP_002533.
FT   VARIANT     15    15       A -> V (IN BALB/C, C57BL/6, BXSB AND
FT                              MRL).
FT   VARIANT     19    19       Q -> R (IN BALB/C, C57BL/6, BXSB AND
FT                              MRL).
FT   VARIANT     76    76       K -> T (IN BALB/C, C57BL/6 AND MRL).
FT   VARIANT     83    86       KAEP -> ATKTEKDPES (IN BALB/C, C57/BL6,
FT                              BXSB AND MRL).
FT   VARIANT     90    91       PP -> LS (IN BALB/C).
FT   VARIANT     94    94       R -> G (IN BALB/C).
FT   VARIANT    102   102       S -> R (IN BALB/C, C57BL/6, BXSB AND
FT                              MRL).
FT   VARIANT    112   114       PIR -> LIH (IN BXSB).
FT   VARIANT    173   173       E -> G (IN BALB/C, C57BL/6 AND MRL).
FT   VARIANT    179   179       Q -> K (IN BALB/C, C57BL/6 AND MRL).
FT   VARIANT    186   186       K -> E (IN BALB/C, C57BL/6 AND MRL).
FT   VARIANT    190   190       V -> I (IN BALB/C, C57BL/6 AND MRL).
FT   VARIANT    192   192       P -> S (IN BALB/C, C57BL/6 AND MRL).
FT   VARIANT    236   236       E -> K (IN BALB/C, C57BL/6 AND MRL).
FT   VARIANT    241   241       R -> C (IN BALB/C).
FT   VARIANT    244   244       R -> H (IN BALB/C AND MRL).

  Query Match             4.5%;  Score 104;  DB 1;  Length 862;
  Best Local Similarity   20.8%;  Pred. No. 15;
  Matches   55;  Conservative   47;  Mismatches  105;  Indels   58;  Gaps   12;

Qy         32 KKGDTVELTC--TASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQ 89
              : |  | | |    |    ::| || :  :   |   || :  |              |
Db        535 RAGQRVLLQCDFAESNPAEVRFFWKKNGSLVQEGRYLSFGSVSPE------------DS 581

Qy         90 GNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLV----FGLTANSDTHLLQGQSLTLTLES 145
              ||:  :: |          | |    |   :|:|       :: :   |:::|:  ||: ||
Db        582 GNYNCMVNN--------SIGETLSQAWNLQVLYAPRRLRVSISPGDHVMEGKKATLSCES 633

Qy        146 ---PPGSSPSVQCRSPRGKNI-QGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVV 201
                 || |  :   |  |:::   |: | :  ||:| :|:: |          |
Db        634 DANPPISQYTWFDSS--GQDLHSSGQKLRLEPLEVQHTGSYRCKGTNGIGTGE------- 684

Qy        202 LAFQKASSIVYKKE--GEQVEFSFPLAFTVEKLTGSGELW-------WQAERA----SSS 248
                :    :: |  |  |::|        |:  |     :|       |:  |:      :
```

```
Db        685 -SPPSTLTVYYSPETIGKRVALGLGFCLTICILA----IWGMKIQKKWKQNRSQQGLQEN 739

Qy        249 KSWITFDLKNKEVSVKRVTQDPKLQ 273
                |  :| ::||:     ::: |: |
Db        740 SSGQSFFVRNKKARRTPLSEGPQSQ 764


RESULT 91
NRCA_CHICK
ID   NRCA_CHICK     STANDARD;      PRT;  1284 AA.
AC   P35331;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neuronal cell adhesion molecule precursor (Nr-CAM) (NgCAM-related cell
DE   adhesion molecule) (Ng-CAM-related) (hBravo).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 25-52; 178-184 AND 581-594.
RC   STRAIN=White leghorn; TISSUE=Embryonic brain;
RX   MEDLINE=91258407; PubMed=2045418;
RA   Grumet M., Mauro V., Burgoon M.P., Edelman G.M., Cunningham B.A.;
RT   "Structure of a new nervous system glycoprotein, Nr-CAM, and its
RT   relationship to subgroups of neural cell adhesion molecules.";
RL   J. Cell Biol. 113:1399-1412(1991).
RN   [2]
RP   SEQUENCE OF 25-1284 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Embryonic brain, and Retina;
RX   MEDLINE=92381110; PubMed=1512296;
RA   Kayyem J.F., Roman J.M., de la Rosa E.J., Schwarz U., Dreyer W.J.;
RT   "Bravo/Nr-CAM is closely related to the cell adhesion molecules L1
RT   and Ng-CAM and has a similar heterodimer structure.";
RL   J. Cell Biol. 118:1259-1270(1992).
CC   -!- FUNCTION: This protein is a cell adhesion molecule involved in
CC       neuron-neuron adhesion, neurite fasciculation, outgrowth of
CC       neurites, etc. Specifically involved in the development of optic
CC       fibres in the retina.
CC   -!- SUBUNIT: Heterodimer of an alpha and a beta chain.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=5;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=P35331-1; Sequence=Displayed;
CC       Name=2; Synonyms=AS10;
CC         IsoId=P35331-2; Sequence=VSP_002603;
CC       Name=3; Synonyms=AS12;
CC         IsoId=P35331-3; Sequence=VSP_002604;
CC       Name=4; Synonyms=AS93;
CC         IsoId=P35331-4; Sequence=VSP_002605;
CC       Name=5; Synonyms=AS-CYT2;
CC         IsoId=P35331-5; Sequence=VSP_002606;
CC   -!- TISSUE SPECIFICITY: Retina and developing brain.
CC   -!- DEVELOPMENTAL STAGE: Expressed in developing neural retina and
CC       embryonic brain tissue.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily.
CC       L1/neurofascin/NgCAM family.
CC   -!- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 5 fibronectin type III domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X58482; CAA41391.1; -.
DR   EMBL; L08960; AAA48632.1; -.
DR   HSSP; P20241; 1CFB.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00060; FN3; 3.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   Immunoglobulin domain; Glycoprotein; Signal; Cell adhesion; Repeat;
KW   Transmembrane; Alternative splicing.
FT   SIGNAL        1     24
FT   CHAIN        25   1284       NEURONAL CELL ADHESION MOLECULE.
FT   DOMAIN       25   1143       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1144   1166       POTENTIAL.
FT   DOMAIN     1167   1284       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       41    129       IG-LIKE C2-TYPE 1.
FT   DOMAIN      136    230       IG-LIKE C2-TYPE 2.
FT   DOMAIN      243    332       IG-LIKE C2-TYPE 3.
FT   DOMAIN      337    424       IG-LIKE C2-TYPE 4.
FT   DOMAIN      430    517       IG-LIKE C2-TYPE 5.
FT   DOMAIN      521    608       IG-LIKE C2-TYPE 6.
FT   DOMAIN      638    699       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      738    799       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      837    906       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      943   1006       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     1057   1114       FIBRONECTIN TYPE-III 5.
FT   DISULFID     63    118       POTENTIAL.
FT   DISULFID    162    213       POTENTIAL.
FT   DISULFID    268    316       POTENTIAL.
FT   DISULFID    358    408       POTENTIAL.
FT   DISULFID    452    501       POTENTIAL.
FT   DISULFID    543    592       POTENTIAL.
FT   CARBOHYD     78     78       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    218    218       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    290    290       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    409    409       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    483    483       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    581    581       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    595    595       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    692    692       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    778    778       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    834    834       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    885    885       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    969    969       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    985    985       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    995    995       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1048   1048       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1059   1059       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1091   1091       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    612    621       Missing (in isoform 2).
FT                                /FTId=VSP_002603.
FT   VARSPLIC   1027   1038       Missing (in isoform 3).
FT                                /FTId=VSP_002604.
FT   VARSPLIC   1039   1131       Missing (in isoform 4).
FT                                /FTId=VSP_002605.
FT   VARSPLIC   1202   1205       Missing (in isoform 5).
FT                                /FTId=VSP_002606.
FT   CONFLICT    209    209       V -> E (IN REF. 2).
FT   CONFLICT    680    680       H -> Q (IN REF. 2).
SQ   SEQUENCE   1284 AA;  141851 MW;  A3570BF9C3D47A0F CRC64;

  Query Match             4.5%;  Score 104;  DB 1;  Length 1284;
  Best Local Similarity   20.0%;  Pred. No. 25;
  Matches   91;  Conservative   68;  Mismatches  172;  Indels  124;  Gaps   23;

Qy         18 LLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLN 77
              | |  :  ||| |  :|: : | | |:   :    |                      |
Db        247 LTPMGSTSNKVEL--RGNVLLLECIAAGLPTPVIRW------------------IKEG 284
```

```
Qy          78 DRADSRRSLWDQGNFPLIIKNLKIEDSDT--YICEVEDQKEEVQLLVFGLTANSDTHLLQ 135
Db         285 GELPANRTFFE--NFKKTLKIIDVSEADSGNYKCTARN-----------TLGSTHHVI- 329

Qy         136 GQSLTL----------TLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTC 184
Db         330 --SVTVKAAPYWITAPRNLVLSPGEDGTLICRA------NGNPKPSISWL---TNGVPIA 378

Qy         185 TVLQNQKKVEFKID---IVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEK--------LT 233
Db         379 IAPEDPSR---KVDGDTIIFSAVQERSSAVYQCNASN-EYGYLLANAFVNVLAEPPRILT 434

Qy         234 GSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLH----LTLPQALP 289
Db         435 PANKLYQVIADSPALIDCAYFGSPKPEIEWFRGVKGSILR-GNEYVFHDNGTLEIPVA-- 491

Qy         290 QYAGSGNLTLALEAKTGKLHQEVNLVVMRATQL-------------QKNLTCEV-WGPTS 335
Db         492 QKDSTGTYTCVARNKLGKTQNEVQLEVKDPTMIIKQPQYKVIQRSAQASFECVIKHDPTL 551

Qy         336 PKLMLSLKLENKEAKVSKR----EKPVWVLN---PEAGMWQCLLSDSGQVLLESNIKVLP 388
Db         552 IPTVIWLK-DNNELPDDERFLVGKDNLTIMNVTDKDDGTYTCIVNTTLDSVSASAV---- 606

Qy         389 TWSTPVHPRASALPAPPTGSALPDPQTASALPDPP 423
Db         607 ---------LTVVAAPPT------PAIIYARPNPP 626


RESULT 92
VGR2_COTJA
ID   VGR2_COTJA     STANDARD;      PRT;  1348 AA.
AC   P52583;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Vascular endothelial growth factor receptor 2 precursor (EC 2.7.1.112)
DE   (VEGFR-2) (Endothelial kinase receptor EK1) (Quek 1) (Quek1).
GN   KDR OR FLK-1 OR EK1.
OS   Coturnix coturnix japonica (Japanese quail).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Coturnix.
OX   NCBI_TaxID=93934;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=97017121; PubMed=8863722;
RA   Eichmann A., Marcelle C., Breant C., Le Douarin N.M.;
RT   "Molecular cloning of Quek 1 and 2, two quail vascular endothelial
RT   growth factor (VEGF) receptor-like molecules.";
RL   Gene 174:3-8(1996).
RN   [2]
RP   SEQUENCE OF 910-1348 FROM N.A.
RC   TISSUE=Spinal cord;
RX   MEDLINE=93378866; PubMed=8396413;
RA   Eichmann A., Marcelle C., Breant C., Le Douarin N.M.;
RT   "Two molecules related to the VEGF receptor are expressed in early
RT   endothelial cells during avian embryonic development.";
RL   Mech. Dev. 42:33-48(1993).
RN   [3]
RP   SEQUENCE OF 764-880 FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=Embryo;
RX   MEDLINE=95301109; PubMed=7781909;
RA   Flamme I., Breier G., Risau W.;
RT   "Vascular endothelial growth factor (VEGF) and VEGF receptor 2 (flk-1)
RT   are expressed during vasculogenesis and vascular differentiation in
RT   the quail embryo.";
RL   Dev. Biol. 169:699-712(1995).
CC   -!- FUNCTION: RECEPTOR FOR VEGF OR VEGF-C. HAS A TYROSINE-PROTEIN
CC       KINASE ACTIVITY. THE VEGF-KINASE LIGAND/RECEPTOR SIGNALING SYSTEM
CC       PLAYS A KEY ROLE IN VASCULAR DEVELOPMENT AND REGULATION OF
CC       VASCULAR PERMEABILITY.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: IN ALL ENDOTHELIAL TISSUES DURING ONSET OF
CC       VASCULARIZATION. IN LATER DEVELOPMENT, PRESENT IN LUNG, HEART,
CC       INTESTINE AND SKIN.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE MESODERM AT ONSET OF
CC       GASTRULATION. FROM DAY 2, CONFINED TO ENDOTHELIAL TISSUES AND
CC       EXPRESSION CONTINUES TO BE WIDESPREAD THROUGHOUT VASCULARIZATION
CC       UNTIL E9 WHERE IT BECOMES RESTRICTED TO SPECIFIC REGIONS SUCH AS
CC       THE SPINAL CHORD AND HEART VALVES.
CC   -!- INDUCTION: IN VITRO, VEGR IS INDUCED BY BASIC FIBROBLAST GROWTH
CC       FACTOR (FGF), UNIQUELY IN THE FIRST 24 H OF CELL CULTURE.
CC   -!- SIMILARITY: Belongs to the Tyr family of protein kinases.
CC       CSF-1/PDGF receptor subfamily.
CC   -!- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X83288; CAA58268.1; -.
DR   EMBL; S65205; AAB28127.1; -.
DR   EMBL; S78345; AAB34594.1; -.
DR   PIR; JC4953; S51656.
DR   HSSP; P11362; 1FGK.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001824; RecepttyrkinsIII.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   Pfam; PF00047; ig; 6.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 2.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 5.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
KW   Angiogenesis; Signal; Transferase; Tyrosine-protein kinase; Receptor;
KW   Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;
KW   Immunoglobulin domain; Repeat.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   1348       VASCULAR ENDOTHELIAL GROWTH FACTOR
FT                                RECEPTOR 2.
FT   DOMAIN       21    756       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    757    777       POTENTIAL.
FT   DOMAIN      778   1348       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       43    106       IG-LIKE C2-TYPE 1.
FT   DOMAIN      138    202       IG-LIKE C2-TYPE 2.
FT   DOMAIN      220    312       IG-LIKE C2-TYPE 3.
FT   DOMAIN      320    405       IG-LIKE C2-TYPE 4.
FT   DOMAIN      412    534       IG-LIKE C2-TYPE 5.
FT   DOMAIN      540    651       IG-LIKE C2-TYPE 6.
FT   DOMAIN      658    744       IG-LIKE C2-TYPE 7.
FT   DOMAIN      825   1155       PROTEIN KINASE.
FT   NP_BIND     831    839       ATP (BY SIMILARITY).
FT   BINDING     859    859       ATP (BY SIMILARITY).
FT   ACT_SITE   1021   1021       BY SIMILARITY.
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     47     47       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     63     63       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     93     93       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    138    138       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    153    153       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    201    201       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    240    240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    290    290       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    310    310       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    365    365       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    386    386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    613    613       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    622    622       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    666    666       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    688    688       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    710    710       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    865    865       A -> S (IN REF. 3).
SQ   SEQUENCE   1348 AA;  150305 MW;  A5E4194A76FD5FB3 CRC64;

  Query Match             4.5%;  Score 104;  DB 1;  Length 1348;
  Best Local Similarity   19.2%;  Pred. No. 27;
  Matches   73;  Conservative   55;  Mismatches  130;  Indels  122;  Gaps   19;

Qy          9 HLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHW-KNSNQIKILGNQGS 67
              |:   |:: | .| :    |       |   | ||| :    :  | | |  :    :|
Db        536 HVTRGLEINLQPRSQLTEK-------DNTSLQCTADKFTFEKLSWYKLSTHV----SQTP 584

Qy         68 FLTKG--PSKLNDRADSRRSL--------WDQGNFPLIIKNLKIEDSDTYICEVEDQKEE 117
              |   |  |  :    |: : |         :     ||::|: ::|   |:|  :|:| :
Db        585 F---GGLPMPVCKNLDALQKLNATVSNVNGENVTLELILRNISLQDGGDYVCIAQDKKAK 641

Qy        118 VQ-LLVFGLTANSDTH-----LLQGQSLTLTLESPPGSSPSVQCRSPRG---------KN 162
              |  ||  ||    |       |: |:  :      | :  | | :   |          ||
Db        642 TQHCLVKHLTVQEPLHPRLVGNLENQTTNI------GETIEVLC-TVNGVPPPNITWFKN 694

Qy        163 IQ------------GGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSI 210
               :            | |||:: ::  :| | :||                 :|  :|| :
Db        695 SETLFEDSGIVLKDGNKTLTIRRVRKEDGGLYTCLACN------------ILGCKKAEAF 742

Qy        211 VYKKEGEQVEFSFPLAFTVEKLTGSGELWW----------QAERASSSKSWITFDLKNKE 260
               :  :| : : :  |   |      :   |           :|        :::  :   |
Db        743 -FSVQGAEEKTNLELIILVGTAVIAMFFWLLLVIILRTVKRANGGDMKTGYLSIIMDPDE 801

Qy        261 VSV----KRVTQD--------PKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKL 308
              | :    :|: |         :|::|| |                | |     :||
Db        802 VPIDEHCERLPYDASKWEFPRDRLKLGKPL-------------GRGAFGQVIEAD---- 843

Qy        309 HQEVNLVVMRATQLQKNLTC 328
                        |  : |  ||
Db        844 ----------AFGIDKTATC 853


RESULT 93
NEO1_HUMAN
ID   NEO1_HUMAN     STANDARD;      PRT;  1461 AA.
AC   Q92859; O00340;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neogenin precursor.
GN   NEO1 OR NGN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97236653; PubMed=9121761;
RA   Meyerhardt J.A., Look A.T., Bigner S.H., Fearon E.R.;
RT   "Identification and characterization of neogenin, a DCC-related
RT   gene.";
RL   Oncogene 14:1129-1136(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97312699; PubMed=9169140;
RA   Vielmetter J., Chen X.-N., Miskevich F., Lane R.P., Yamakawa K.,
RA   Korenberg J.R., Dreyer W.J.;
RT   "Molecular characterization of human neogenin, a DCC-related protein,
RT   and the mapping of its gene (NEO1) to chromosomal position 15q22.3-
RT   q23.";
RL   Genomics 41:414-421(1997).
CC   -!- FUNCTION: May be involved as a regulatory protein in the
CC       transition of undifferentiated proliferating cells to their
CC       differentiated state. May also function as a cell adhesion
CC       molecule in a broad spectrum of embryonic and adult tissues.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=Q92859-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q92859-2; Sequence=VSP_002593;
CC   -!- TISSUE SPECIFICITY: Widely expressed and also in cancer cell
CC       lines.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.
CC   -!- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.
CC   -!- SIMILARITY: Contains 6 fibronectin type III domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U61262; AAB17263.1; -.
DR   EMBL; U72391; AAC51287.1; -.
DR   HSSP; P02751; 1TTF.
DR   Genew; HGNC:7754; NEO1.
DR   MIM; 601907; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 4.
KW   Cell adhesion; Repeat; Signal; Transmembrane; Immunoglobulin domain;
KW   Glycoprotein; Alternative splicing.
FT   SIGNAL        1     33       POTENTIAL.
FT   CHAIN        34   1461       NEOGENIN.
FT   DOMAIN       34   1105       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1106   1126       POTENTIAL.
FT   DOMAIN     1127   1461       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       52    141       IG-LIKE C2-TYPE 1.
FT   DOMAIN      152    238       IG-LIKE C2-TYPE 2.
FT   DOMAIN      243    336       IG-LIKE C2-TYPE 3.
FT   DOMAIN      341    426       IG-LIKE C2-TYPE 4.
FT   DOMAIN      436    533       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      536    629       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      630    729       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      735    829       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      850    950       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      951   1052       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     1118   1121       POLY-VAL.
FT   DISULFID     74    129       BY SIMILARITY.
FT   DISULFID    173    221       BY SIMILARITY.
FT   DISULFID    270    320       BY SIMILARITY.
FT   DISULFID    362    410       BY SIMILARITY.
FT   CARBOHYD     73     73       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    210    210       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    326    326       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    639    639       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    715    715       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    909    909       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1248   1300       Missing (in isoform 2).
FT                                /FTId=VSP_002593.
FT   CONFLICT    168    168       G -> N (IN REF. 2).
SQ   SEQUENCE   1461 AA;  159958 MW;  7AAE897E69635A21 CRC64;

  Query Match             4.5%;  Score 104;  DB 1;  Length 1461;
  Best Local Similarity   22.8%;  Pred. No. 30;
  Matches   91;  Conservative   45;  Mismatches  158;  Indels  106;  Gaps   21;

Qy         30 LGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQ 89
              |  :| :| | |:|  : | :  ||           |:|       ||  :| || |
Db         63 LSVRGSSVILNCSAYSEPSPKIEWK---------KDGTF-------LNLVSDDRRQLLPD 106

Qy         90 GNFPLIIKNL-----KIEDSDTYIC--EVEDQ----KEEVQLLVFGL----TANSDTHLL 134
              |:  | | |:         |   | |   ||          :|:| ||    :    : :
Db        107 GS--LFISNVVHSKHNKPDEGYYQCVATVESLGTIISRTAKLIVAGLPRFTSQPEPSSVY 164

Qy        135 QGQSLTLTLESPPGSSPSV---QCRSP-----RGKNIQGGKTLSVSQLELQDSGTWTCTV 186
              |    | |       | |   | | |     |   :  | | :|     | | : | |
Db        165 AGNGAILNCEVNADLVPFVRWEQNRQPLLLDDRVIKLPSG-MLVISNATEGDGGLYRCVV 223

Qy        187 LQN-----QKKVEFKI-----DIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSG 236
                  :|| |:       |  | | |  | : :  |: |     |       |
Db        224 ESGGPPKYSDEVELKVLPDPEVISDLVFLKQPSPLVRVIGQDV----VLPCVASGLPTPT 279

Qy        237 ELWWQAERA---SSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAG 293
                | : | |    ||:  :   |    : :  ||:|     |    :           |
Db        280 IKWMKNEEALDTESSERLVL--LAGGSLEISDVTED---DAGTYFCI----------AD 323

Qy        294 SGNLT------LALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENK 347
              :|| |      | ::|:   | |  |:    :  :     ||| |  :| :
Db        324 NGNETIEAQAELTVQAQPEFLKQPTNIYAHESMDIV--FECEVTGKPTPTVK-------- 373

Qy        348 EAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
                      || | :     : ||  ::: | |::||
Db        374 ----------WVKNGD----MVIPSDYFKIVKEHNLQVL 398


RESULT 94
CENF_HUMAN
ID   CENF_HUMAN     STANDARD;      PRT;  3210 AA.
AC   P49454; Q13171; Q13246;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   CENP-F kinetochore protein (Centromere protein F) (Mitosin) (AH
DE   antigen).
GN   CENPF.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast carcinoma;
RX   MEDLINE=95348175; PubMed=7542657;
RA   Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
RT   "CENP-F is a protein of the nuclear matrix that assembles onto
RT   kinetochores at late G2 and is rapidly degraded after mitosis.";
RL   J. Cell Biol. 130:507-518(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95379848; PubMed=7651420;
RA   Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
RA   Jones D., Yang-Feng T.L., Lee W.-H.;
RT   "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RT   that is specifically involved in mitotic-phase progression.";
RL   Mol. Cell. Biol. 15:5017-5029(1995).
RN   [3]
RP   SEQUENCE OF 2194-3210 FROM N.A.
RX   MEDLINE=95336446; PubMed=7612011;
RA   Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT   "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
RT   domain sufficient for nuclear localization.";
RL   Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=95370296; PubMed=7642639;
RA   Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT   "The C terminus of mitosin is essential for its nuclear localization,
RT   centromere/kinetochore targeting, and dimerization.";
RL   J. Biol. Chem. 270:19545-19550(1995).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=98437347; PubMed=9763420;
RA   Chan G.K.T., Schaar B.T., Yen T.J.;
RT   "Characterization of the kinetochore binding domain of CENP-E reveals
RT   interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL   J. Cell Biol. 143:49-63(1998).
RN   [6]
RP   FARNESYLATION.
RX   MEDLINE=20459117; PubMed=10852915;
RA   Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA   Bishop W.R., Kirschmeier P.;
RT   "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT   and CENP-F and alter the association of CENP-E with the
RT   microtubules.";
RL   J. Biol. Chem. 275:30451-30457(2000).
CC   -!- FUNCTION: Probably required for kinetochore function, involved in
CC       chromosome segregation during mitosis. Interacts with
CC       retinoblastoma protein (RB), CENP-E and BUBR1.
CC   -!- SUBUNIT: Homo- or heterodimer.
CC   -!- SUBCELLULAR LOCATION: Nuclear matrix (but not in the nucleolus),
CC       reorganization to the kinetochore/centromere (coronal surface of
CC       the outer plate) and the spindle during mitosis.
CC   -!- DEVELOPMENTAL STAGE: Gradually accumulates during the cell cycle.
CC   -!- PTM: Hyperphosphorylated during mitosis.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19769; AAA82889.1; -.
DR   EMBL; U30872; AAA82935.1; -.
DR   EMBL; U25725; AAA86889.1; -.
DR   PIR; PC4035; PC4035.
DR   Genew; HGNC:1857; CENPF.
DR   GK; P49454; -.
DR   MIM; 600236; -.
DR   GO; GO:0005699; C:kinetochore; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0005819; C:spindle; TAS.
DR   GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR   GO; GO:0007088; P:regulation of mitosis; TAS.
KW   Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
KW   Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat; Polymorphism;
KW   Lipoprotein; Prenylation.
FT   DOMAIN       14    197       COILED COIL (POTENTIAL).
FT   DOMAIN      273    769       COILED COIL (POTENTIAL).
FT   DOMAIN      823   1328       COILED COIL (POTENTIAL).
FT   DOMAIN     1642   1746       COILED COIL (POTENTIAL).
FT   DOMAIN     1862   2987       COILED COIL (POTENTIAL).
FT   DOMAIN     2207   2568       2 X 177 AA TANDEM REPEATS.
FT   REPEAT     2207   2386       1.
FT   REPEAT     2389   2568       2.
```

```
FT   DOMAIN      3015   3032       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   LIPID       3207   3207       S-farnesyl cysteine.
FT   VARIANT     3202   3202       K -> N (in dbSNP:7289).
FT                                 /FTId=VAR_014839.
FT   CONFLICT      16     16       T -> A (IN REF. 2).
FT   CONFLICT     250    250       L -> Q (IN REF. 2).
FT   CONFLICT     272    272       G -> D (IN REF. 2).
FT   CONFLICT     611    611       MISSING (IN REF. 2).
FT   CONFLICT    1494   1589       MISSING (IN REF. 2).
FT   CONFLICT    1611   1611       V -> A (IN REF. 2).
FT   CONFLICT    1811   1811       V -> L (IN REF. 2).
FT   CONFLICT    2242   2243       ER -> DG (IN REF. 3).
FT   CONFLICT    2335   2335       L -> Q (IN REF. 3).
FT   CONFLICT    2492   2492       D -> N (IN REF. 2).
FT   CONFLICT    2545   2561       ELNERVAALHNDQEACK -> SSMREWQPCIMTKKPVS
FT                                 (IN REF. 3).
SQ   SEQUENCE   3210 AA;  367589 MW;  11D83324960E4334 CRC64;

  Query Match             4.5%;  Score 104;  DB 1;  Length 3210;
  Best Local Similarity   20.0%;  Pred. No. 83;
  Matches   84;  Conservative   58;  Mismatches  133;  Indels  144;  Gaps   17;

Qy         46 KKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSD 105
              ||:   : |  ||:|              | |:||       ||    |:||: |
Db       2867 KKTKMDNLKYVNQLK-------------KENERA--------QGKMKLLIKSCK----- 2899

Qy        106 TYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQG 165
                ::|::|| :|  :  | |       | :  | |:        : : :  :
Db       2900 ----QLEEEKEILQKELSQLQA------AQEKQKTGTVMDTKVDELTTEIKELK------ 2943

Qy        166 GKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKE-------GEQ 218
               :||      |  :     |::| : :|:|   :::      |   |::       |
Db       2944 -ETLEEKTKEADEYLDKYCSLLISHEKLEKAKEMLETQVAHLCSQQSKQDSRGSPLLGPV 3002

Qy        219 VEFSFPLAFTVEKLTGSGELWWQAERASSSKSW-----------ITFDLKNKEVSVKRV- 266
              |    |:    ||   ||:     :|  ||  |          :|  |:|:  :   :
Db       3003 VPGPSPIPSVTEKRLSSGQNKASGKRQRSSGIWENGGGPTPATPESFSKKSKKAVMSGIH 3062

Qy        267 ----TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQL 322
                |:  : :           |: ||:    |      :  |||     :  ::| | :
Db       3063 PAEDTEGTEFE-----------PEGLPEVVKKG----FADIPTGK----TSPYILRRTTM 3103

Qy        323 QKNLTCEVWGPTSPKLM--------LSLKLEN--------------KEAKVSKREKPVWV 360
                      |||:|          |||  ||              :: ||::|  ||
Db       3104 ATR--------TSPRLAAQKLALSPLSLGKENLAESSKPTAGGSRSQKVKVAQR-SPV-- 3152

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH------------PRASALPAPPTG 407
                         ||| :| |   | :|  : |               |   :|:|  |
Db       3153 ------------DSGTILREPTTKSVPVNNLPERSPTDSPREGLRVKRGRLVPSPKAG 3198


RESULT 95
KV5S_MOUSE
ID   KV5S_MOUSE     STANDARD;      PRT;   108 AA.
AC   P01652;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-V region J606.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=82099361; PubMed=6798111;
RA   Johnson N., Slankard J., Paul L., Hood L.;
RT   "The complete V domain amino acid sequences of two myeloma inulin-
RT   binding proteins.";
RL   J. Immunol. 128:302-307(1982).
CC   -!- MISCELLANEOUS: THIS CHAIN WAS ISOLATED FROM MYELOMA PROTEINS THAT
CC       BIND BETA(2-1)-FRUCTOFURANOSYL MOIETIES (INULIN).
DR   PIR; A92811; KVMS06.
DR   HSSP; P01607; 1REI.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region.
FT   DOMAIN        1     23       FRAMEWORK-1.
FT   DOMAIN       24     34       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       35     49       FRAMEWORK-2.
FT   DOMAIN       50     56       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       57     88       FRAMEWORK-3.
FT   DOMAIN       89     97       COMPLEMENTARITY-DETERMINING-3.
FT   DOMAIN       98    108       FRAMEWORK-4.
FT   DISULFID     23     88       BY SIMILARITY.
FT   NON_TER     108    108
SQ   SEQUENCE   108 AA;  11810 MW;  8DE4DD31076F2AFB CRC64;

  Query Match             4.5%;  Score 103.5;  DB 1;  Length 108;
  Best Local Similarity   39.0%;  Pred. No. 1.1;
  Matches   30;  Conservative    5;  Mismatches   37;  Indels    5;  Gaps    2;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQG-NF 92
              || | :|| |||  ||  :|        |    | | | | |  || |    | :|
Db         16 GDIVTMTCQASQGTSINLNWFQQKP----GKAPKLLIYGASNLEDGVPSRFSGSRYGTDF 71

Qy         93 PLIIKNLKIEDSDTYIC 109
                | | :|: ||  || |
Db         72 TLTISSLEDEDMATYFC 88


RESULT 96
KV1W_HUMAN
ID   KV1W_HUMAN     STANDARD;      PRT;   129 AA.
AC   P04431;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Ig kappa chain V-I region Walker precursor.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85014148; PubMed=6091049;
RA   Klobeck H.G., Combriato G., Zachau H.G.;
RT   "Immunoglobulin genes of the kappa light chain type from two human
RT   lymphoid cell lines are closely related.";
RL   Nucleic Acids Res. 12:6995-7006(1984).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X00965; CAA25477.1; ALT_TERM.
DR   PIR; A01883; K1HUWK.
DR   HSSP; P01607; 1REI.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0003823; F:antigen binding; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Immunoglobulin V region; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23    129       IG KAPPA CHAIN V-I REGION WALKER.
```

```
FT   DOMAIN       23     45       FRAMEWORK-1.
FT   DOMAIN       46     56       COMPLEMENTARITY-DETERMINING-1.
FT   DOMAIN       57     71       FRAMEWORK-2.
FT   DOMAIN       72     78       COMPLEMENTARITY-DETERMINING-2.
FT   DOMAIN       79    110       FRAMEWORK-3.
FT   DOMAIN      111    119       COMPLEMENTARITY-DETERMINING-3.
FT   DOMAIN      120    129       FRAMEWORK-4.
FT   DISULFID     45    110       BY SIMILARITY.
FT   NON_TER     129    129
SQ   SEQUENCE   129 AA;  14069 MW;  F941FA07D4AFC2F9 CRC64;

  Query Match             4.5%;  Score 103.5;  DB 1;  Length 129;
  Best Local Similarity   32.8%;  Pred. No. 1.4;
  Matches   43;  Conservative   10;  Mismatches   61;  Indels   17;  Gaps    5;

Qy          1 MNRGVPFRHLLLVLQLALLPAA------TQGNKVVLGKKGDTVELTCTASQKKSIQFHWK 54
              |:  || :  || | |  |  |       ||     :    || | :|| |||  |   :|
Db          1 MDMRVPAQ--LLGLLLLWLRGARCDIQMTQSPSSLSASVGDRVTITCRASQSISNYLNWY 58

Qy         55 NSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQG-NFPLIIKNLKIEDSDTYICEVED 113
                  |    |   | |     || |    | :| | | :|: ||| || |
Db         59 QQKP----GKAPKLLIYAASSLQSGVTSRFSGSGSGTDFTLTISSLQPEDSATYYC---- 110

Qy        114 QKEEVQLLVFG 124
              |:    |: ||
Db        111 QQSYSTLITFG 121


RESULT 97
NTRI_MOUSE
ID   NTRI_MOUSE     STANDARD;      PRT;   344 AA.
AC   Q99PJ0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurotrimin precursor.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=Brain;
RA   Kim T.H., Choi S.C., Kim J., Jeon J.W., Kim K.D., Lee S.H.;
RT   "Cloning and expression of mouse neurotrimin gene in the developing
RT   nervous system.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Neural cell adhesion molecule.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC       family.
CC   -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF282980; AAK00276.1; -.
DR   EMBL; BC023307; AAH23307.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal; Lipoprotein.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    321       NEUROTRIMIN.
FT   PROPEP      322    344       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       39    126       IG-LIKE C2-TYPE 1.
FT   DOMAIN      136    218       IG-LIKE C2-TYPE 2.
FT   DOMAIN      222    309       IG-LIKE C2-TYPE 3.
FT   DISULFID     57    115       POTENTIAL.
FT   DISULFID    157    201       POTENTIAL.
FT   DISULFID    243    295       POTENTIAL.
FT   CARBOHYD     44     44       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     70     70       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    284    284       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    321    321       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       321    321       GPI-anchor amidated asparagine
FT                                (Potential).
FT   CONFLICT     75     75       L -> P (IN REF. 1).
FT   CONFLICT     92     92       S -> G (IN REF. 1).
FT   CONFLICT    119    119       T -> I (IN REF. 1).
FT   CONFLICT    187    187       E -> Q (IN REF. 1).
FT   CONFLICT    213    213       R -> P (IN REF. 1).
FT   CONFLICT    225    225       I -> F (IN REF. 1).
SQ   SEQUENCE   344 AA;  37984 MW;  C885BBA52C148554 CRC64;

  Query Match             4.5%;  Score 103.5;  DB 1;  Length 344;
  Best Local Similarity   20.3%;  Pred. No. 5;
  Matches   75;  Conservative   55;  Mismatches  166;  Indels   73;  Gaps   13;

Qy         11 LLVLQLALLPAATQGNKVVLG----------KKGDTVELTCTASQKKSIQFHWKNSN 57
              |:|: | ||     |  |  |            ::|::  | ||   : : :  | | :
Db         14 LVVVSLRLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDNRVT-RVAWLNRS 72

Qy         58 QIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEE 117
                |  ||    |      |::       . |  : : |:|: : |   | | |:      .
Db         73 TILYAGNDKWCLDPRVVLLSN---------TQTQYSIEIQNVDVYDEGPYTCSVQTDNHP 123

Qy        118 --------VQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCR--SPRGKN-IQGG 166
                     ||:    :  :|| : :| :::||  :     |:|  |  ||:    :
Db        124 KTSRVHLIVQVSPKIVEISSDISINEGNNISLTCIATGRPEPTVTWRHISPKAVGFVSED 183

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLA 226
              : | :  :  : || : |:  :                  |: :| :   :|  ::|
Db        184 EYLEIQGITREQSGEYECSASND---------------VAAPVVRR---VKVTVNYPPY 224

Qy        227 FTVEKLTG-----SGELWWQAERASSSK-SWITFDLKNKEVSVKRVTQDPKLQMGKKLPL 280
              :  | ||       | |  :|    |::  |   |         ||: :  |    :  |
Db        225 ISEAKGTGVPVGQKGTLQCEASAVPSAEFQWFKDD--------KRLVEGKKGVKVENRPF 276
```

```
Qy        281 HLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCE----VWGPTSP 336
Db        277 LSKLTFFNVSEHDYGNYTCVASNKLGHTNASIMLFGPGAVSEVNNGTSRRAGCIW--LLP 334

Qy        337 KLMLSLKLE 345
Db        335 LLVLHLLLK 343


RESULT 98
NTRI_RAT
ID   NTRI_RAT       STANDARD;      PRT;   344 AA.
AC   Q62718;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Neurotrimin precursor (GP65).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 217-229.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=95198094; PubMed=7891157;
RA   Struyk A.F., Canoll P.D., Wolfgang M.J., Rosen C.L., D'Eustachio P.,
RA   Salzer J.L.;
RT   "Cloning of neurotrimin defines a new subfamily of differentially
RT   expressed neural cell adhesion molecules.";
RL   J. Neurosci. 15:2141-2156(1995).
CC   -!- FUNCTION: Neural cell adhesion molecule.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: Central nervous system.
CC   -!- DEVELOPMENTAL STAGE: Expressed at high levels in several
CC       developing projection systems: in neurons of the thalamus,
CC       subplate, and lower cortical laminae in the forebrain and in the
CC       pontine nucleus, cerebellar granule cells, and Purkinje cells in
CC       the hindbrain.
CC   -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLON
CC       family.
CC   -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U16845; AAA67445.1; -.
DR   PIR; I56551; I56551.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal; Lipoprotein.
FT   SIGNAL        1     31       POTENTIAL.
FT   CHAIN        32    321       NEUROTRIMIN.
FT   PROPEP      322    344       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       39    126       IG-LIKE C2-TYPE 1.
FT   DOMAIN      136    218       IG-LIKE C2-TYPE 2.
FT   DOMAIN      222    309       IG-LIKE C2-TYPE 3.
FT   DISULFID     57    115       POTENTIAL.
FT   DISULFID    157    201       POTENTIAL.
FT   DISULFID    243    295       POTENTIAL.
FT   CARBOHYD     44     44       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     70     70       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    152    152       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    216    216       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    284    284       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    321    321       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   LIPID       321    321       GPI-anchor amidated asparagine
FT                                (Potential).
SQ   SEQUENCE   344 AA;  37998 MW;  CBB39BE53B33B224 CRC64;

  Query Match             4.5%;  Score 103.5;  DB 1;  Length 344;
  Best Local Similarity   20.1%;  Pred. No. 5;
  Matches   74;  Conservative   54;  Mismatches  168;  Indels   73;  Gaps   13;

Qy         11 LLVLQLALLPAATQGNKVVLG------------KKGDTVELTCTASQKKSIQFHWKNSN 57
Db         14 LVVVSLRLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDNRVT-RVAWLNRS 72

Qy         58 QIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEE 117
Db         73 TILYAGNDKWCLDPRVVLLSN---------TQTQYSIEIQNVDVYDEGPYTCSVQTDNHP 123

Qy        118 --------VQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCR--SPRGKN-IQGG 166
Db        124 KTSRVHLIVQVSPKIVEISSDISINEGNNISLTCIATGRPEPTVTWRHISPKAVGFVSED 183

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLA 226
Db        184 EYLEIQGITREQSGEYECSASN---------DVAAPVVRRVN----------VTVNYPPY 224

Qy        227 FTVEKLTG-----SGELWWQAERASSSK-SWITFDLKNKEVSVKRVTQDPKLQMGKKLPL 280
Db        225 ISEAKGTGVPVGQKGTLQCEASAVPSAEFQWFKDD--------KRLVEGKKGVKVENRPF 276

Qy        281 HLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCE----VWGPTSP 336
Db        277 LSRLTFFNVSEHDYGNYTCVASNKLGHTNASIMLFGPGAVSEVNNGTSRRAGCIW--LLP 334

Qy        337 KLMLSLKLE 345
Db        335 LLVLHLLLK 343


RESULT 99
GAG_SIVAT
ID   GAG_SIVAT      STANDARD;      PRT;   519 AA.
AC   P05892;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   GAG polyprotein [Contains: Core proteins P17, P24, and P15].
GN   GAG.
OS   Simian immunodeficiency virus (TYO-1 isolate) (SIV-AGM).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11731;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88232906; PubMed=3374586;
RA   Fukasawa M., Miura T., Hasegawa A., Morikawa S., Tsujimoto H.,
RA   Miki K., Kitamura T., Hayami M.;
RT   "Sequence of simian immunodeficiency virus from African green monkey,
RT   a new member of the HIV/SIV group.";
RL   Nature 333:457-461(1988).
CC   -!- MISCELLANEOUS: This is an African green monkey isolate.
CC   -!- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X07805; CAA30657.1; -.
DR   HSSP; P05888; 1AAF.
```

```
DR   HIV; X07805; GAG$AGMTY.
DR   InterPro; IPR000721; Gag_p24.
DR   InterPro; IPR008916; Retrov_capsid_C.
DR   InterPro; IPR008919; Retrov_capsid_N.
DR   InterPro; IPR000071; Retrovir_p17.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00540; Gag_p17; 1.
DR   Pfam; PF00607; Gag_p24; 1.
DR   Pfam; PF00098; zf-CCHC; 2.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00234; HIV1MATRIX.
DR   SMART; SM00343; ZnF_C2HC; 2.
DR   PROSITE; PS50158; ZF_CCHC; 2.
KW   AIDS; Core protein; Polyprotein; Zinc-finger; Repeat.
FT   CHAIN         1    141       CORE PROTEIN P17.
FT   CHAIN       142    519       CORE PROTEINS P24 AND P15.
FT   ZN_FING     397    414       CCHC-TYPE 1.
FT   ZN_FING     418    435       CCHC-TYPE 2.
SQ   SEQUENCE   519 AA;  58143 MW;  85A3AC06BCCDCA38 CRC64;

  Query Match             4.5%;  Score 103.5;  DB 1;  Length 519;
  Best Local Similarity   19.6%;  Pred. No. 8.5;
  Matches  101;  Conservative   62;  Mismatches  162;  Indels  191;  Gaps   26;

Qy         18 LLPAATQGNKVVLG------------KKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQ 65
              | |  ::| | :               | || |   |  |   :    |:: :
Db         67 LEPTGSEGLKSLFNLVCVLYCLHKEQKVKDTEEAVATVRQHCHLVEKEKSATE------- 119

Qy         66 GSFLTKGPSKLNDR------------ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVE 112
                 |   | ||:              |  : : |   : ||  : |     : ::  ||
Db        120 ----TSSGQKKNDKGIAAPPGCSQNFPAQQQGNAW--VHVPLSPRTL-----NAWVKAVE 168

Qy        113 DQK--EEV----QLLVFGLTANSDTHLL------------------QGQSLTLTLESPP 147
              ::|   |:    | |  | |      :|                 :     :|   |
Db        169 EKKFGAEIVPMFQALSEGCTPYDINQMLNVLGDHQGALQIVKEIINEEAAQWDVTHPLPA 228

Qy        148 GSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDI-------V 200
              |  |: | | ||| :| |  :    |||       |  |          ::|:       :
Db        229 GPLPAGQLRDPRGSDIAGTTSSVQEQLE------WIYT-------ANPRVDVGAIYRRWI 275

Qy        201 VLAFQKA------SSIVYKKEGEQVEFS-----FPLAFTVEKLTGSGELWWQAERASSSK 249
              :|  ||       |::  ::| :  |       |  |   |:   |||:          |
Db        276 ILGLQKCVKMYNPVSVLDIRQGPKEPFKDYVDRFYKAIRAEQ--ASGEV----------K 323

Qy        250 SWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLH 309
               |:|     : : ::    | |: : | | :| || : |    | |          |  :
Db        324 QWMT-----ESLLIQNANPDCKVIL-KGLGMHPTLEEMLTACQGVG----------GPSY 367

Qy        310 QEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREK-PVWVLN------ 362
              :   :  | |   :|:  :  ||                    ||:: |:   |
Db        368 KAKVMAEMMQTMQNQNMV-QQGGP------------------KRQRPPLRCYNCGKFGH 407

Qy        363 -----PEAGMWQCLLSDS---------GQV-------LLESNIKVLPTWSTPVHPRASAL 401
                  ||    :||             |||       : :  :  |  :    | |
Db        408 MQRQCPEPRKTKCLKCGKLGHLAKDCRGQVNFLGYGRWMGAKPRNFPAATLGAEPSA--- 464

Qy        402 PAPPTGSALPDP-------------QTASALPDPPA 424
              | ||:|:   ||            |      :|||
Db        465 PPPPSGTTPYDPAKKLLQQYAEKGKQLREQKRNPPA 500


RESULT 100
PEC1_HUMAN
ID   PEC1_HUMAN     STANDARD;      PRT;   738 AA.
AC   P16284;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Platelet endothelial cell adhesion molecule precursor (PECAM-1)
DE   (CD31 antigen) (EndoCAM) (GPIIA').
GN   PECAM1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90193682; PubMed=1690453;
RA   Newman P.J., Berndt M.C., Gorski J., White J.C. II, Lyman S.,
RA   Paddock C., Muller W.A.;
RT   "PECAM-1 (CD31) cloning and relation to adhesion molecules of the
RT   immunoglobulin gene superfamily.";
RL   Science 247:1219-1222(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91340830; PubMed=1874786;
RA   Albelda S.M., Muller W.A., Buck C.A., Newman P.J.;
RT   "Molecular and cellular properties of PECAM-1 (endoCAM/CD31): a novel
RT   vascular cell-cell adhesion molecule.";
RL   J. Cell Biol. 114:1059-1068(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91060975; PubMed=1700999;
RA   Stockinger H., Gadd S.J., Eher R., Majdic O., Kasinrek W.,
RA   Schreiber W., Strass B., Schnabl E., Knapp W.;
RT   "Molecular characterization and functional analysis of the leukocyte
RT   surface protein CD31.";
RL   J. Immunol. 145:3889-3897(1990).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95086194; PubMed=7994021;
RA   Kirschbaum N.E., Gumina R.J., Newman P.J.;
RT   "Organization of the gene for human platelet/endothelial cell
RT   adhesion molecule-1 shows alternatively spliced isoforms and a
RT   functionally complex cytoplasmic domain.";
RL   Blood 84:4028-4037(1994).
RN   [5]
RP   VARIANT VAL-125.
RX   MEDLINE=96133740; PubMed=8532023;
RA   Behar E., Chao N.J., Hiraki D.D., Krishnaswamy S., Brown B.W.,
RA   Zehnder J.L., Grumet F.C.;
RT   "Polymorphism of adhesion molecule CD31 and its role in acute
RT   graft-versus-host disease.";
RL   New Engl. J. Med. 334:286-291(1996).
RN   [6]
RP   VARIANT VAL-125.
RX   MEDLINE=21649465; PubMed=11791967;
RA   Casals-Pascual C., Allen S., Allen A., Kai O., Lowe B., Pain A.,
RA   Roberts D.J.;
RT   "Codon 125 polymorphism of CD31 and susceptibility to malaria.";
RL   Am. J. Trop. Med. Hyg. 65:736-737(2001).
CC   -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE EXPRESSED ON
CC       PLATELETS AND AT ENDOTHELIAL CELL INTERCELLULAR JUNCTIONS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- PTM: PHOSPHORYLATED ON SER AND TYR RESIDUES AFTER CELLULAR
CC       ACTIVATION.
CC   -!- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD31 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd31.htm".
CC   -!- DATABASE: NAME=R&D Systems' cytokine source book: PECAM1;
CC       WWW="http://www.rndsystems.com/asp/g_sitebuilder.asp?bodyId=218".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M28526; AAA36429.1; -.
DR   EMBL; M37780; AAA36186.1; -.
DR   EMBL; L34657; AAA60057.1; -.
DR   EMBL; L34631; AAA60057.1; JOINED.
DR   EMBL; L34637; AAA60057.1; JOINED.
DR   EMBL; L34638; AAA60057.1; JOINED.
```

```
DR   EMBL; L34639; AAA60057.1; JOINED.
DR   EMBL; L34640; AAA60057.1; JOINED.
DR   EMBL; L34641; AAA60057.1; JOINED.
DR   EMBL; L34642; AAA60057.1; JOINED.
DR   EMBL; L34644; AAA60057.1; JOINED.
DR   EMBL; L34645; AAA60057.1; JOINED.
DR   EMBL; L34649; AAA60057.1; JOINED.
DR   EMBL; L34655; AAA60057.1; JOINED.
DR   PIR; A40096; A40096.
DR   Genew; HGNC:8823; PECAM1.
DR   MIM; 173445; -.
DR   GO; GO:0005886; C:plasma membrane; TAS.
DR   GO; GO:0005888; C:proteoglycan integral to plasma membrane; TAS.
DR   GO; GO:0008037; P:cell recognition; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003599; Ig.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00409; IG; 3.
DR   PROSITE; PS50835; IG_LIKE; 4.
KW   Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
KW   Repeat; Phosphorylation; Signal; Polymorphism.
FT   SIGNAL        1     27
FT   CHAIN        28    738       PLATELET ENDOTHELIAL CELL ADHESION
FT                                 MOLECULE.
FT   DOMAIN        28    601       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     602    620       POTENTIAL.
FT   DOMAIN       621    738       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        35    121       IG-LIKE C2-TYPE 1.
FT   DOMAIN       145    233       IG-LIKE C2-TYPE 2.
FT   DOMAIN       236    315       IG-LIKE C2-TYPE 3.
FT   DOMAIN       328    401       IG-LIKE C2-TYPE 4.
FT   DOMAIN       424    493       IG-LIKE C2-TYPE 5.
FT   DOMAIN       499    591       IG-LIKE C2-TYPE 6.
FT   DISULFID      57    109       POTENTIAL.
FT   DISULFID     152    206       POTENTIAL.
FT   DISULFID     256    304       POTENTIAL.
FT   DISULFID     347    386       POTENTIAL.
FT   DISULFID     431    476       POTENTIAL.
FT   DISULFID     523    572       POTENTIAL.
FT   CARBOHYD      52     52       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD      84     84       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     301    301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     320    320       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     344    344       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     453    453       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     551    551       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   MOD_RES      713    713       PHOSPHORYLATION (POTENTIAL).
FT   VARIANT      125    125       L -> V (in dbSNP:668).
FT                                 /FTId=VAR_013145.
FT   CONFLICT     563    563       S -> N (IN REF. 3 AND 4).
FT   CONFLICT     670    670       R -> G (IN REF. 3).
SQ   SEQUENCE   738 AA;  82536 MW;  C57BBFA200A407A6 CRC64;

  Query Match             4.5%;  Score 103.5;  DB 1;  Length 738;
  Best Local Similarity   25.8%;  Pred. No. 13;
  Matches   49;  Conservative   30;  Mismatches   84;  Indels   27;  Gaps    8;

Qy        101 IEDSDTYICEVEDQK-EEVQLLVFGLT--------ANSDTHLLQGQSLTLTLESPPGSSP 151
              :| |  | |:||  :  :|  :|  :|         :| ||| ||: | |:  | ||: |
Db        296 VEHSGNYTCKVESSRISKVSSIVVNITELFSKPELESSFTHLDQGERLNLSC-SIPGAPP 354

Qy        152 SVQCRSPRGKNIQGGKTLSVSQLE-------LQDSGTWTCTV-LQNQKKVEFKIDIVVLA 203
              :          ||   |: ||| :          ||||: ||  :    |    : |||
Db        355 A-------NFTIQKEDTI-VSQTQDFTKIASKSDSGTYICTAGIDKVVKKSNTVQIVVCE 406

Qy        204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
                    | |  : | ::    :    | ::|:  : :|  : |      | :  :  |
Db        407 MLSQPRISYDAQFEVIK-GQTIEVRCESISGTLPISYQLLKTSKVLENSTKNSNDPAVFK 465

Qy        264 KRVTQDPKLQ 273

                    |:| : |
Db        466 DNPTEDVEYQ 475


Search completed: August  3, 2004, 13:09:14
Job time : 14.2211 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:51:44 ; Search time 31.1057 Seconds
(without alignments)
4594.975 Million cell updates/sec

Title: SEQ5
Perfect score: 2325
Sequence: 1 MNRGVPFRHLLLVLQLALLP.....VISFLLGLGLGVACVLARTR 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1715	73.8	397	6	O09261	O09261 cercopithec
2	1713	73.7	397	6	O09260	O09260 cercopithec
3	1711	73.6	397	6	Q95NE9	Q95ne9 cercopithec
4	1708	73.5	397	6	O09259	O09259 cercopithec
5	1706	73.4	397	6	O09263	O09263 cercopithec
6	1699	73.1	397	6	O09262	O09262 cercopithec
7	1699	73.1	397	6	O02805	O02805 cercopithec
8	1615.5	69.5	457	6	Q8HZT8	Q8hzt8 callithrix
9	1585.5	68.2	457	6	Q8HZT7	Q8hzt7 saimiri sci
10	1188	51.1	455	6	Q9XS78	Q9xs78 delphinapte
11	1087	46.8	474	6	P79355	P79355 felis silve
12	985	42.4	457	11	Q61396	Q61396 mus musculu
13	954.5	41.1	433	11	O55054	O55054 mus musculu
14	400	17.2	86	6	O77596	O77596 mandrillus
15	400	17.2	86	6	O77597	O77597 mandrillus
16	397	17.1	86	6	O77594	O77594 cercopithec
17	397	17.1	86	6	O77599	O77599 theropithec
18	395	17.0	86	6	O77595	O77595 cercocebus
19	392	16.9	86	6	O77598	O77598 papio sp. (
20	388	16.7	86	6	O77601	O77601 lophocebus
21	383	16.5	86	6	O77600	O77600 lophocebus
22	357	15.4	71	4	Q13969	Q13969 homo sapien
23	305.5	13.1	99	6	Q29027	Q29027 sus scrofa
24	280.5	12.1	99	6	Q29028	Q29028 sus scrofa
25	278.5	12.0	120	11	O88650	O88650 marmota mon
26	271.5	11.7	482	13	Q90WB5	Q90wb5 anas platyr
27	245	10.5	487	13	Q9W6V7	Q9w6v7 gallus gall
28	237	10.2	90	11	P70443	P70443 mus musculu
29	174	7.5	739	6	Q28260	Q28260 canis famil
30	165	7.1	650	6	Q9GKR2	Q9gkr2 bos taurus
31	165	7.1	739	6	Q9GKR3	Q9gkr3 bos taurus
32	162	7.0	739	11	Q63669	Q63669 rattus norv
33	161.5	6.9	538	6	Q29123	Q29123 sus scrofa
34	157	6.8	955	4	Q8NFP4	Q8nfp4 homo sapien
35	155.5	6.7	538	6	Q28939	Q28939 sus scrofa
36	152	6.5	31	4	Q9UDE5	Q9ude5 homo sapien
37	152	6.5	1496	4	Q92626	Q92626 homo sapien
38	151.5	6.5	739	6	Q865F2	Q865f2 oryctolagus
39	145	6.2	1431	11	Q80U60	Q80u60 mus musculu
40	144.5	6.2	437	4	Q8IZP8	Q8izp8 homo sapien
41	144	6.2	760	11	Q91X98	Q91x98 mus musculu
42	143	6.2	5636	4	Q96RW7	Q96rw7 homo sapien
43	142	6.1	739	11	Q8K0X1	Q8k0x1 mus musculu
44	142	6.1	885	6	Q8HYV1	Q8hyv1 sus scrofa
45	142	6.1	886	6	Q8HYV2	Q8hyv2 sus scrofa
46	141.5	6.1	397	11	Q9JIE0	Q9jie0 mus musculu
47	140.5	6.0	702	4	Q8N4D0	Q8n4d0 homo sapien
48	140.5	6.0	17352	5	Q95YM2	Q95ym2 procambarus
49	139.5	6.0	403	5	Q9VP08	Q9vp08 drosophila
50	138	5.9	398	11	Q07763	Q07763 mus musculu
51	138	5.9	2828	4	Q9NR99	Q9nr99 homo sapien
52	137.5	5.9	1259	11	Q9QY38	Q9qy38 mus musculu
53	136.5	5.9	1395	5	O44924	O44924 drosophila
54	136.5	5.9	1465	11	Q7TQG5	Q7tqg5 mus musculu
55	136.5	5.9	4117	5	Q8IRV9	Q8irv9 drosophila
56	136.5	5.9	4179	5	Q9W4Y4	Q9w4y4 drosophila
57	136.5	5.9	4223	5	Q8MPN3	Q8mpn3 drosophila
58	136.5	5.9	4228	5	Q8IRV8	Q8irv8 drosophila
59	136	5.8	512	4	Q96DN8	Q96dn8 homo sapien
60	136	5.8	545	5	Q9VCT4	Q9vct4 drosophila
61	134.5	5.8	1035	5	Q9NEG1	Q9neg1 drosophila
62	134	5.8	487	13	Q7T2H2	Q7t2h2 gallus gall
63	134	5.8	1445	11	Q63155	Q63155 rattus norv
64	133.5	5.7	6620	4	Q96AA2	Q96aa2 homo sapien
65	133	5.7	2693	5	Q8ISF3	Q8isf3 caenorhabdi
66	133	5.7	2708	5	Q8ISF4	Q8isf4 caenorhabdi
67	133	5.7	2780	5	Q8MNS0	Q8mns0 caenorhabdi
68	133	5.7	2808	5	Q8MNS1	Q8mns1 caenorhabdi
69	133	5.7	18519	5	Q8ISF6	Q8isf6 caenorhabdi
70	133	5.7	18534	5	Q8ISF7	Q8isf7 caenorhabdi
71	132.5	5.7	352	13	Q9W6V2	Q9w6v2 gallus gall
72	132.5	5.7	1596	4	Q9HCL6	Q9hcl6 homo sapien
73	132	5.7	811	5	Q9VNP2	Q9vnp2 drosophila
74	132	5.7	1011	5	Q24273	Q24273 drosophila
75	132	5.7	1248	6	Q9XT41	Q9xt41 cercopithec
76	132	5.7	1255	6	Q7YQL7	Q7yql7 pongo pygma
77	132	5.7	2212	4	Q8NHN3	Q8nhn3 homo sapien
78	132	5.7	6831	5	Q23550	Q23550 caenorhabdi
79	132	5.7	7158	5	Q23551	Q23551 caenorhabdi
80	131.5	5.7	848	5	Q25198	Q25198 hydra atten
81	131	5.6	483	13	Q7SX76	Q7sx76 brachydanio
82	131	5.6	827	11	Q80TU8	Q80tu8 mus musculu
83	131	5.6	1255	4	Q7Z3Z9	Q7z3z9 homo sapien
84	131	5.6	1255	6	Q7YQL8	Q7yql8 pan troglod
85	131	5.6	1340	4	Q8NDA2	Q8nda2 homo sapien
86	131	5.6	5175	5	Q8I0L3	Q8i0l3 caenorhabdi
87	131	5.6	5198	5	O76518	O76518 caenorhabdi
88	130	5.6	4162	13	Q98918	Q98918 gallus gall
89	129.5	5.6	340	11	O88654	O88654 mus musculu

90	129	5.5	287	11	Q9D7B8	Q9d7b8 mus musculu
91	128.5	5.5	564	13	Q7ZU00	Q7zu00 brachydanio
92	128.5	5.5	1395	5	Q9W213	Q9w213 drosophila
93	128.5	5.5	1413	11	Q8CGB3	Q8cgb3 mus musculu
94	127	5.5	2629	11	O08653	O08653 rattus norv
95	126.5	5.4	399	11	Q9ESE5	Q9ese5 mus musculu
96	126.5	5.4	484	5	Q26475	Q26475 schistocerc
97	126	5.4	785	11	Q7TNP4	Q7tnp4 mus musculu
98	126	5.4	3197	5	Q9W1D5	Q9w1d5 drosophila
99	125.5	5.4	513	4	O00481	O00481 homo sapien
100	125	5.4	605	11	Q921P2	Q921p2 mus musculu
101	125	5.4	838	11	Q8C4B2	Q8c4b2 mus musculu
102	125	5.4	838	11	Q8BQ96	Q8bq96 mus musculu
103	125	5.4	1056	13	Q7ZW34	Q7zw34 brachydanio
104	125	5.4	1746	4	Q8WY19	Q8wy19 homo sapien
105	125	5.4	6875	6	Q28733	Q28733 oryctolagus
106	124.5	5.4	355	4	Q7Z3B1	Q7z3b1 homo sapien
107	124.5	5.4	358	13	Q8UV81	Q8uv81 brachydanio
108	124.5	5.4	536	11	Q8BJE2	Q8bje2 mus musculu
109	124.5	5.4	544	13	Q7ZZ85	Q7zz85 brachydanio
110	124	5.3	520	11	Q925P2	Q925p2 mus musculu
111	124	5.3	542	4	Q8NHN5	Q8nhn5 homo sapien
112	124	5.3	731	6	Q8SPI6	Q8spi6 macropus eu
113	124	5.3	758	6	Q9N2H7	Q9n2h7 sus scrofa
114	124	5.3	7962	4	Q10465	Q10465 homo sapien
115	124	5.3	34350	4	Q8WZ42	Q8wz42 homo sapien
116	123.5	5.3	325	13	Q8UV52	Q8uv52 brachydanio
117	123	5.3	18412	13	Q7ZZ61	Q7zz61 brachydanio
118	122.5	5.3	338	4	Q8IV49	Q8iv49 homo sapien
119	122.5	5.3	495	4	Q9HCY1	Q9hcy1 homo sapien
120	122.5	5.3	733	6	Q8SQ83	Q8sq83 trichosurus
121	122.5	5.3	2013	11	Q9ERC8	Q9erc8 mus musculu
122	122	5.2	859	15	O92937	O92937 human immun
123	122	5.2	1323	13	Q08476	Q08476 gallus gall
124	122	5.2	2013	11	Q8VHZ8	Q8vhz8 rattus norv
125	121.5	5.2	577	11	Q9D2Z1	Q9d2z1 mus musculu

ALIGNMENTS

```
RESULT 1
O09261
ID   O09261      PRELIMINARY;      PRT;   397 AA.
AC   O09261;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4 (Fragment).
GN   CD4.
OS   Cercopithecus sabaeus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=60711;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98017879; PubMed=9379478;
RA   Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA   Corbet S., Barre-Sinoussi F., Allan J.S.;
RT   "Relation between phylogeny of African green monkey CD4 genes and
RT   their respective simian immunodeficiency virus genes.";
RL   J. Med. Primatol. 26:120-128(1997).
DR   EMBL; AF001225; AAB60872.1; -.
DR   HSSP; P01730; 1WIQ.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER       1      1
FT   NON_TER     397    397
SQ   SEQUENCE   397 AA;  43945 MW;  80C963B92A868CD3 CRC64;

  Query Match             73.8%;  Score 1715;  DB 6;  Length 397;
  Best Local Similarity   89.5%;  Pred. No. 1.9e-123;
  Matches  332;  Conservative   15;  Mismatches   24;  Indels    0;  Gaps    0;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              |||||||||||||| ||| : |||||||||||||| |||||||| ||| || |||:|||
Db          1 VVLGKKGDTVELTCNASQNTTTQFHWKNSNQIKILGKQGSFLTKGSSKLRDRIDSRKSLW 60

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPP 147
              ||| | :|||||||||:|||||||::||||:||||||||||||||||||||||||||||
Db         61 DQGCFSMIIKNLKIEDSETYICEVENKKEEVELLVFGLTANSDTHLLQGQSLTLTLESPP 120

Qy        148 GSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKA 207
              ||||||:|||||||||||:|||| ||| |||||||||| |:|  ||||||||||||||
Db        121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERQDSGTWTCTVSQDQNMVEFKIDIVVLAFQKA 180

Qy        208 SSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVT 267
              || |||||||||||||||:|||||||||||||||||||||||||||||||||||:||
Db        181 SSTVYKKEGEQVEFSFPLAFTLEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKQVT 240

Qy        268 QDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLT 327
              ||||||||||||:|||||||||||||||||||||||||||||||||||||||| |:|||
Db        241 QDPKLQMGKKLPLNLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQFQENLT 300

Qy        328 CEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
              |||||||||||||||||| | |||: | ||||||||||||||||||||||||||||||
Db        301 CEVWGPTSPKLMLSLKLENKAATVSKQAKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 360

Qy        388 PTWSTPVHPRA 398
              ||| ||| | |
Db        361 PTWPTPVQPMA 371


RESULT 2
O09260
ID   O09260      PRELIMINARY;      PRT;   397 AA.
AC   O09260;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4 (Fragment).
GN   CD4.
OS   Cercopithecus sabaeus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=60711;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98017879; PubMed=9379478;
RA   Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA   Corbet S., Barre-Sinoussi F., Allan J.S.;
RT   "Relation between phylogeny of African green monkey CD4 genes and
RT   their respective simian immunodeficiency virus genes.";
RL   J. Med. Primatol. 26:120-128(1997).
DR   EMBL; AF001224; AAB60871.1; -.
DR   HSSP; P01730; 1WIQ.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER       1      1
FT   NON_TER     397    397
```

```
SQ   SEQUENCE   397 AA;  43882 MW;  478BB277E992EE89 CRC64;

  Query Match             73.7%;  Score 1713;  DB 6;  Length 397;
  Best Local Similarity   80.2%;  Pred. No. 2.7e-123;
  Matches  341;  Conservative   17;  Mismatches   37;  Indels   30;  Gaps    2;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              ||||||||||||||| |||| : |||||||||||||| |||||||| ||| || |||:|||
Db          1 VVLGKKGDTVELTCNASQKTTTQFHWKNSNQIKILGKQGSFLTKGSSKLRDRIDSRKSLW 60

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPP 147
              ||| | :|||||||||:|||||||:::|||:||||||||||||||||||||||||||||
Db         61 DQGCFSMIIKNLKIEDSETYICEVENKEEEVELLVFGLTANSDTHLLQGQSLTLTLESPP 120

Qy        148 GSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKA 207
              |||||||:|||||||||||:|||| ||| |||||||| | |:|  ||||||||||||||
Db        121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERQDSGTWTCNVSQDQNTVEFKIDIVVLAFQKA 180

Qy        208 SSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVT 267
              || ||||||||||||||||||:||||||||||||||||||||||||||||||||||:||
Db        181 SSTVYKKEGEQVEFSFPLAFTLEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKQVT 240

Qy        268 QDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLT 327
              ||||||||||||:|||||||||||||||||||||||||||||||||||||| |:|||
Db        241 QDPKLQMGKKLPLNLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQFQENLT 300

Qy        328 CEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
              |||||||||||||||||| | |||: | ||||||||||||||||||||||||||||||
Db        301 CEVWGPTSPKLMLSLKLENKAATVSKQAKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 360

Qy        388 PTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGV-A 446
              ||| ||| |                              ||      |   ||  |||:
Db        361 PTWPTPVQP----------------------------MALIVLGGVAGLLLFTGLGIFL 391

Qy        447 CVLAR 451
              ||  |
Db        392 CVRCR 396


RESULT 3
Q95NE9
ID   Q95NE9      PRELIMINARY;      PRT;   397 AA.
AC   Q95NE9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4 (Fragment).
GN   CD4.
OS   Cercopithecus pygerythrus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=60710;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98017879; PubMed=9379478;
RA   Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA   Corbet S., Barre-Sinoussi F., Allan J.S.;
RT   "Relation between phylogeny of African green monkey CD4 genes and
RT   their respective simian immunodeficiency virus genes.";
RL   J. Med. Primatol. 26:120-128(1997).
DR   EMBL; AF001227; AAB60874.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER       1      1
FT   NON_TER     397    397
```

```
SQ   SEQUENCE   397 AA;  43946 MW;  21C3E30882ABFBC0 CRC64;

  Query Match             73.6%;  Score 1711;  DB 6;  Length 397;
  Best Local Similarity   80.2%;  Pred. No. 3.8e-123;
  Matches  341;  Conservative   16;  Mismatches   38;  Indels   30;  Gaps    2;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              |||||||||||||| |||  : |||||||||||||| |||||||| ||| || |||:|||
Db          1 VVLGKKGDTVELTCNASQNTTTQFHWKNSNQIKILGKQGSFLTKGSSKLRDRIDSRKSLW 60

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPP 147
              ||| | :|||||||||:|||||||::||||:||||||||||||||||||||||||||||
Db         61 DQGCFSMIIKNLKIEDSETYICEVENKKEEVELLVFGLTANSDTHLLQGQSLTLTLESPP 120

Qy        148 GSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKA 207
              ||||||:|||||||||||:|||| ||| |||||||| |:|  ||||||||||||||
Db        121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERQDSGTWTCTVSQDQNTVEFKIDIVVLAFQKA 180

Qy        208 SSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVT 267
              || ||||||||||||||||||:||||||||||||||||||||||||||||||||||:||
Db        181 SSTVYKKEGEQVEFSFPLAFTLEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKQVT 240

Qy        268 QDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLT 327
              ||||||||||||:|||||||||||||||||||||||||||||||||||||| |:|||
Db        241 QDPKLQMGKKLPLNLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQFQENLT 300

Qy        328 CEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
              |||||||||||||||||| | |||: | ||||||| ||||||||||||||||||||
Db        301 CEVWGPTSPKLMLSLKLENKAATVSKQAKAVWVLNPEEGMWQCLLSDSGQVLLESNIKVL 360

Qy        388 PTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGV-A 446
              ||| ||| |                              ||      |   ||  |||:
Db        361 PTWPTPVQP----------------------------MALIVLGGVAGLLLFTGLGIFF 391

Qy        447 CVLAR 451
              ||  |
Db        392 CVRCR 396


RESULT 4
O09259
ID   O09259      PRELIMINARY;      PRT;   397 AA.
AC   O09259;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4 (Fragment).
GN   CD4.
OS   Cercopithecus sabaeus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=60711;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98017879; PubMed=9379478;
RA   Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA   Corbet S., Barre-Sinoussi F., Allan J.S.;
RT   "Relation between phylogeny of African green monkey CD4 genes and
RT   their respective simian immunodeficiency virus genes.";
RL   J. Med. Primatol. 26:120-128(1997).
DR   EMBL; AF001223; AAB60870.1; -.
DR   HSSP; P01730; 1WIQ.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER       1      1
```

```
FT   NON_TER     397    397
SQ   SEQUENCE   397 AA;  43881 MW;  7CE39AD0F8506C81 CRC64;

  Query Match             73.5%;  Score 1708;  DB 6;  Length 397;
  Best Local Similarity   80.0%;  Pred. No. 6.5e-123;
  Matches  340;  Conservative   18;  Mismatches   37;  Indels   30;  Gaps    2;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              |||||||||||||| |||| : |||||||||||||| |||||||| ||| || |||:|||
Db          1 VVLGKKGDTVELTCNASQKTTTQFHWKNSNQIKILGKQGSFLTKGSSKLRDRIDSRKSLW 60

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPP 147
              ||| | :|||||||||||:|||||||:::|||:||||||||||||||||||||||||||||
Db         61 DQGCFSMIIKNLKIEDSETYICEVENKEEEVELLVFGLTANSDTHLLQGQSLTLTLESPP 120

Qy        148 GSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKA 207
              ||||||:|||||||||||:|||| ||| |||||||| | |:|  ||||||||||||||||
Db        121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERQDSGTWTCNVSQDQNTVEFKIDIVVLAFQKA 180

Qy        208 SSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVT 267
              || ||||||||||||||||:||||||||||||||||||||:|||||||:||
Db        181 SSTVYKKEGEQVEFSFPLAFTLEKLTGSGELWWQAERASSSKSWITFNLKNKEVSVKQVT 240

Qy        268 QDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLT 327
              |||||||||||||:|||||||||||||||||||||||||||||||||||| |:|||
Db        241 QDPKLQMGKKLPLNLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQFQENLT 300

Qy        328 CEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
              |||||||||||||||||| | |||: | ||||||||||||||||||||||||||||||||
Db        301 CEVWGPTSPKLMLSLKLENKAATVSKQAKAVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 360

Qy        388 PTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGV-A 446
              ||| ||| |                              ||     |   ||  |||:
Db        361 PTWPTPVQP----------------------------MALIVLGGVAGLLLFTGLGIFL 391

Qy        447 CVLAR 451
              ||  |
Db        392 CVRCR 396


RESULT 5
O09263
ID   O09263      PRELIMINARY;      PRT;   397 AA.
AC   O09263;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4 (Fragment).
GN   CD4.
OS   Cercopithecus tantalus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=60712;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98017879; PubMed=9379478;
RA   Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA   Corbet S., Barre-Sinoussi F., Allan J.S.;
RT   "Relation between phylogeny of African green monkey CD4 genes and
RT   their respective simian immunodeficiency virus genes.";
RL   J. Med. Primatol. 26:120-128(1997).
DR   EMBL; AF001222; AAB60869.1; -.
DR   HSSP; P01730; 1WIQ.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER       1      1
FT   NON_TER     397    397
SQ   SEQUENCE   397 AA;  43994 MW;  A3CD031535A51524 CRC64;

  Query Match             73.4%;  Score 1706;  DB 6;  Length 397;
  Best Local Similarity   88.9%;  Pred. No. 9.3e-123;
  Matches  330;  Conservative   15;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              |||||||||||||| |||  : |||||||||||||| |||||||| ||| || |||:|||
Db          1 VVLGKKGDTVELTCNASQNTTTQFHWKNSNQIKILGKQGSFLTKGSSKLRDRIDSRKSLW 60

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPP 147
              ||| | :|||||||||||:|||||||::||||:||||||||||||||||||||||||||||
Db         61 DQGCFSMIIKNLKIEDSETYICEVENKKEEVELLVFGLTANSDTHLLQGQSLTLTLESPP 120

Qy        148 GSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKA 207
              ||||||:|||||||||||:|||| ||| ||||||||| |:|  ||||||||||||||||
Db        121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERQDSGTWTCTVSQDQNTVEFKIDIVVLAFQKA 180

Qy        208 SSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVT 267
              || ||||||||||||||||:||||||||||||||||||||||||||||||||:||
Db        181 SSTVYKKEGEQVEFSFPLAFTLEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKQVT 240

Qy        268 QDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLT 327
              ||||||||||||:|||||||||||||||||||||||||||||||||||| |:|||
Db        241 QDPKLQMGKKLPLNLTLPQALPQYAGSGNFTLALEAKTGKLHQEVNLVVMRATQFQENLT 300

Qy        328 CEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
              |||||||||||||||||| | |||: | ||||||| ||||||||||||||||||||||||
Db        301 CEVWGPTSPKLMLSLKLENKAATVSKQAKAVWVLNPEEGMWQCLLSDSGQVLLESNIKVL 360

Qy        388 PTWSTPVHPRA 398
              ||| ||| | |
Db        361 PTWPTPVQPMA 371


RESULT 6
O09262
ID   O09262      PRELIMINARY;      PRT;   397 AA.
AC   O09262;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4 (Fragment).
GN   CD4.
OS   Cercopithecus tantalus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=60712;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98017879; PubMed=9379478;
RA   Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA   Corbet S., Barre-Sinoussi F., Allan J.S.;
RT   "Relation between phylogeny of African green monkey CD4 genes and
RT   their respective simian immunodeficiency virus genes.";
RL   J. Med. Primatol. 26:120-128(1997).
DR   EMBL; AF001221; AAB60868.1; -.
DR   HSSP; P01730; 1WIQ.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER       1      1
FT   NON_TER     397    397
SQ   SEQUENCE   397 AA;  43954 MW;  CF7F2F5D82335B0D CRC64;
```

```
  Query Match             73.1%;  Score 1699;  DB 6;  Length 397;
  Best Local Similarity   88.9%;  Pred. No. 3.2e-122;
  Matches  330;  Conservative   15;  Mismatches   26;  Indels    0;  Gaps    0;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              ||||||||||||||| |||  : |||||||||||||| |||||||| ||| || |||:|||
Db          1 VVLGKKGDTVELTCNASQNTTTQFHWKNSNQIKILGKQGSPLTKGSSKLRDRIDSRKSLW 60

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPP 147
              ||| | :||||||||||:|||||||::||||:||||||||||||||||||||||||||||
Db         61 DQGCFSMIIKNLKIEDSETYICEVENKKEEVELLVFGLTANSDTHLLQGQSLTLTLESPP 120

Qy        148 GSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKA 207
              ||||||:|||||||||||||:|||| ||| |||||||||| |:|  ||||||||||||||
Db        121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERQDSGTWTCTVSQDQNTVEFKIDIVVLAFQKA 180

Qy        208 SSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVT 267
              || |||||||||||||||||:|||||||| ||||||||||||||||||||||||:||
Db        181 SSTVYKKEGEQVEFSFPLAFTLEKLTGSGELRWQAERASSSKSWITFDLKNKEVSVKQVT 240

Qy        268 QDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLT 327
              |||:|||||||||||||||||||||||||||||||||||||||||||||||||||| |:|||
Db        241 QDPELQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQFQENLT 300

Qy        328 CEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
              |||||||||||||||||| | |||: | ||||||| ||||||||||||||||||||||
Db        301 CEVWGPTSPKLMLSLKLENKAATVSKQAKAVWVLNPEEGMWQCLLSDSGQVLLESNIKVL 360

Qy        388 PTWSTPVHPRA 398
              ||| ||| | |
Db        361 PTWPTPVQPMA 371


RESULT 7
O02805
ID   O02805      PRELIMINARY;      PRT;   397 AA.
AC   O02805; O77593;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE   (Fragment).
GN   CD4.
OS   Cercopithecus aethiops (Green monkey) (Grivet).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=9534;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral blood;
RX   MEDLINE=98017879; PubMed=9379478;
RA   Fomsgaard A., Mueller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA   Corbet S., Barre-Sinoussi F., Allan J.S.;
RT   "Relation between phylogeny of African green monkey CD4 genes and
RT   their respective simian immunodeficiency virus genes.";
RL   J. Med. Primatol. 26:120-128(1997).
RN   [2]
RP   SEQUENCE OF 80-165 FROM N.A.
RX   MEDLINE=98320644; PubMed=9656488;
RA   Harris E.E., Disotell T.R.;
RT   "Nuclear gene trees and the phylogenetic relationships of the
RT   mangabeys (Primates: Papionini).";
RL   Mol. Biol. Evol. 15:892-900(1998).
CC   -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC       RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE V-LIKE AND 3 C2-LIKE DOMAINS.
DR   EMBL; AF001226; AAB60873.1; -.
DR   EMBL; AF057380; AAC25124.1; -.
DR   HSSP; P01730; 1WIQ.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
KW   Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW   Lipoprotein; Palmitate; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       <1    370       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    371    391       POTENTIAL.
FT   DOMAIN      392   >397       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       <1     98       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       99    176       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      177    290       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      291    347       IG-LIKE C2-TYPE DOMAIN 3.
FT   CARBOHYD     15     15       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     30     30       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     14     82       BY SIMILARITY.
FT   DISULFID    128    157       BY SIMILARITY.
FT   DISULFID    301    343       BY SIMILARITY.
FT   LIPID       392    392       S-palmitoyl cysteine (By similarity).
FT   LIPID       395    395       S-palmitoyl cysteine (By similarity).
FT   NON_TER     397    397
SQ   SEQUENCE   397 AA;  43980 MW;  F74C42E22B196155 CRC64;

  Query Match             73.1%;  Score 1699;  DB 6;  Length 397;
  Best Local Similarity   79.8%;  Pred. No. 3.2e-122;
  Matches  339;  Conservative   16;  Mismatches   40;  Indels   30;  Gaps    2;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              ||||||||||||||| |||| : ||||||||| |||| |||||||| ||| || |||:|||
Db          1 VVLGKKGDTVELTCNASQKTTTQFHWKNSNQTKILGKQGSFLTKGSSKLRDRIDSRKSLW 60

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPP 147
              ||| | :||||||||||:|||||||::||||:||||||||||||||||||||||||||||
Db         61 DQGCFSMIIKNLKIEDSETYICEVENKKEEVELLVFGLTANSDTHLLQGQSLTLTLESPP 120

Qy        148 GSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKA 207
              ||||||:|||||||||| |:|||| ||| |||||||||| |:|  ||||||||||||||
Db        121 GSSPSVKCRSPRGKNIQVGRTLSVPQLERQDSGTWTCTVSQDQNTVEFKIDIVVLAFQKA 180

Qy        208 SSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVT 267
              || |||||||||||||||| |:|||||||||||||||||||||||||||||||||:||
Db        181 SSTVYKKEGEQVEFSFPLALTLEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKQVT 240

Qy        268 QDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLT 327
              |||:|||||||||||||||||||||||||||||||||||||||||||||||||| |:|||
Db        241 QDPELQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQFQENLT 300

Qy        328 CEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
              |||||||||||||||||| | |||: | ||||||| ||||||||||||||||||||||
Db        301 CEVWGPTSPKLMLSLKLENKAATVSKQAKAVWVLNPEEGMWQCLLSDSGQVLLESNIKVL 360

Qy        388 PTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGV-A 446
              ||| ||| |                                  ||     |   ||  |||:
Db        361 PTWPTPVQP--------------------------MALIVLGGVAGLLLFTGLGIFF 391

Qy        447 CVLAR 451
              ||  |
Db        392 CVRCR 396


RESULT 8
Q8HZT8
ID   Q8HZT8      PRELIMINARY;      PRT;   457 AA.
AC   Q8HZT8;
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lymphocyte antigen CD4.
OS   Callithrix jacchus (Common marmoset).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callithrix.
OX   NCBI_TaxID=9483;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22174698; PubMed=12186836;
RA   LaBonte J.A., Babcock G.J., Patel T., Sodroski J.;
RT   "Blockade of HIV-1 Infection of New World Monkey Cells Occurs
RT   Primarily at the Stage of Virus Entry.";
RL   J. Exp. Med. 196:431-445(2002).
DR   EMBL; AF452616; AAN14532.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
SQ   SEQUENCE   457 AA;  50878 MW;  718CFDB78D97F59B CRC64;

  Query Match             69.5%;  Score 1615.5;  DB 6;  Length 457;
  Best Local Similarity   70.7%;  Pred. No. 1e-115;
  Matches  324;  Conservative   37;  Mismatches   58;  Indels   39;  Gaps    4;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              || |:|||| |||||||||| | | |  |||||||: ||| |   | || :||||| |||||
Db          1 MNGGIPFRHSLLVLQLALLIAVTHGKTVVLGKKGEMVELPCETSLKKKLQFHWKTSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||| |||||:||| ||| :| ||::| ||:|:||||||:|:::|||:|||||||| :||||:|
Db         61 ILGIQGSFVTKGQSKLANRIDSKQSSWDRGSFPLIIRNVQVEDSETYICEVESKKEEVEL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
               ||||| | |||||||||||||||||||||:| ||||| |:| ||| :||||:||||
Db        121 QVFGLTVNPDTHLLQGQSLTLTLESPPGSSPSVECMSPRGKTIRGMKTLFMSQLEIQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              || ||| |: : | |||:|||||||:||| |||||||||||||||  |:|||||| |
Db        181 TWKCTVSQHLELV-FKINIVVLAFQQASSTVYKKEGEQVEFSFPLAFAAEQLTGSGELCW 239

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||:|||||:| |:|| || ||||||||:|||||| |||||||||||||||| |||
Db        240 QAERASSSESWITFNLTNQEVCVKLVTQDPKLRMGKKLPFHLTLPQALPQYAGSGNFTLA 299

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |: |||||||| |||||||| ||||||||||||:|||||||:||||||| |||
Db        300 LKGKTGKLHQEVKLVVMRATQLQNNLTCEVWGPTSPKLVLSLKLENQEAKVSKREKAVWV 359

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||| | |||||||||||||| ::|||||| ||
Db        360 LNPEVGAWQCLLSDSGQVLLESKVEVLPTWSPPVQ------------------------ 394

Qy        421 DPPAASALPAALAVISFLLGL----GLGV-ACVLARTR 453
                      | || |:  : ||    |||:  ||  | |
Db        395 --------PMALIVLGGVAGLLVFTGLGIFLCVRCRHR 424


RESULT 9
Q8HZT7
ID   Q8HZT7      PRELIMINARY;      PRT;   457 AA.
AC   Q8HZT7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Lymphocyte antigen CD4.
OS   Saimiri sciureus (Common squirrel monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.
OX   NCBI_TaxID=9521;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22174698; PubMed=12186836;
RA   LaBonte J.A., Babcock G.J., Patel T., Sodroski J.;
RT   "Blockade of HIV-1 Infection of New World Monkey Cells Occurs
RT   Primarily at the Stage of Virus Entry.";
RL   J. Exp. Med. 196:431-445(2002).
DR   EMBL; AF452617; AAN14533.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
SQ   SEQUENCE   457 AA;  50899 MW;  B164DA89E70C575A CRC64;

  Query Match             68.2%;  Score 1585.5;  DB 6;  Length 457;
  Best Local Similarity   70.1%;  Pred. No. 2e-113;
  Matches  321;  Conservative   34;  Mismatches   64;  Indels   39;  Gaps    4;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              || |:|||||||||||||||| |  |||||||: ||| |   | ||:: |||| |::||
Db          1 MNGGIPFRHLLLVLQLALLPAATHGKTVVLGKKGEVVELPCETSLKKNVPFHWKTSDRIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKDLKIEDSETYICEVESKKEEVEL 120
              ||| |  |:|:| ||| || ||:|| ||:|:|||:||: :||||:||||||| :||||:|
Db         61 ILGVQNYFVTRGQSKLTDRIDSKRSSWDRGSFPLLIKDARIEDSETYICEVESKKEEVEL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||| ||||||||||||||||||||:| ||||| |:| |||||||| : |||
Db        121 QVFGLTANPDTHLLQGQSLTLTLESPPGSSPSVECTSPRGKRIRGRKTLSVSQLGIPDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              || ||| |: : | |:|:|||||||:||| |||||||||||||||  | |||||| |
Db        181 TWKCTVFQHLELV-FEINIVVLAFQQASSTVYKKEGEQVEFSFPLAFAAETLTGSGELCW 239

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||||:|  :|| || |||||||:|||||||||| |||||||||| |||
Db        240 QAERASSSKSWITFNLTKQEVYVKLVTQDPKLRMGKKLPLHLTLAQALPQYAGSGNFTLA 299

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |: |||||||||||||| |||| ||||||||||||||||||:||||||| |||
Db        300 LKGKTGKLHQEVNLVVMRVTQLQNNLTCEVWGPTSPKLMLSLKLENQEAKVSKREKAVWV 359

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||| | |||||||||||||  : ||| | ||
Db        360 LNPEPGAWQCLLSDSGQVLLESKFEALPTRSPPVQ------------------------ 394

Qy        421 DPPAASALPAALAVISFLLGL----GLGV-ACVLARTR 453
                      | | |:  : ||    |||:  ||  | |
Db        395 --------PMVLIVLGGVAGLLAFTGLGIFLCVRCRHR 424


RESULT 10
Q9XS78
ID   Q9XS78      PRELIMINARY;      PRT;   455 AA.
AC   Q9XS78;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4.
OS   Delphinapterus leucas (Beluga whale).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
OC   Monodontidae; Delphinapterus.
OX   NCBI_TaxID=9749;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Thymus;
RX   MEDLINE=99216435; PubMed=10199913;
RA   Romano T.A., Ridgway S.H., Felten D.L., Quaranta V.;
RT   "Molecular cloning and characterization of CD4 in an aquatic mammal,
RT   the white whale Delphinapterus leucas.";
RL   Immunogenetics 49:376-383(1999).
DR   EMBL; AF071799; AAD23738.1; -.
DR   HSSP; P01730; 1WIQ.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 3.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 2.
SQ   SEQUENCE   455 AA;  50499 MW;  AA532FD4411AA5D1 CRC64;

  Query Match             51.1%;  Score 1188;  DB 6;  Length 455;
  Best Local Similarity   56.4%;  Pred. No. 7e-83;
  Matches  257;  Conservative   48;  Mismatches  105;  Indels   46;  Gaps    6;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |:     ||| ||||| :||| ||| |||||| |:  || | ||| ||: | |||| | |
Db          1 MDPRTSLRHLFLVLQLVMLPAGTQGKKVVLGKAGELAELPCKASQNKSLFFSWKNSYQTK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||  | |  || | |: | :|: :|||||:|||:||:|:: || |||||||||:| ||:|
Db         61 ILGRHGYFWHKGASNLHSRVESKINLWDQGSFPLVIKDLEVPDSGTYICEVEDKKIEVEL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
               || |||:||| || |||||||||| | ||:|||| : |  |     |:||: |: |||||
Db        121 QVFRLTASSDTRLLLGQSLTLTLEGPSGSNPSVQWKGPGNKRKNEAKSLSLPQVGLQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGEL-W 239
              |||||| | |: : |   |:|||||: || || |||||: ||||| |  | |  |||| |
Db        181 TWTCTVSQAQQTLVFNKHILVLAFQEVSSTVYAKEGEQMNFSFPLTFGDENL--SGELSW 238

Qy        240 WQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTL 299
              ||:  || :||||| | | :|:| :  :| ||:| | ||||||||||||||||||||||
Db        239 LQAKGNSSPESWITFKLNNGKVTVGKARKDLKLRMSKALPLHLTLPQALPQYAGSGNLTL 298

Qy        300 ALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVW 359
              |    |||:|||||||| |:   :||||| |||||:|:|||| ||:  :|| ::| |
Db        299 NL--TKGKLYQEVNLVVMRVTKSPNSLTCEVLGPTSPRLILSLKKENQSMRVSDQQKLVT 356

Qy        360 VLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASAL 419
              || |||||||||||| |:||||| :|:||
Db        357 VLGPEAGMWQCLLSDKGKVLLESKVKILP------------------------------ 385

Qy        420 PDPPAASALPAALAVI-----SFLLGLGLGVACVLA 450
                |  | | |  |||:     | ||  |    |: :
Db        386 --PVLAHAWPKLLAVVLGGITSLLLLAGF---CIFS 416


RESULT 11
P79355
ID   P79355      PRELIMINARY;      PRT;   474 AA.
AC   P79355;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4 antigen precursor.
OS   Felis silvestris catus (Cat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX   NCBI_TaxID=9685;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Miyazawa T.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92165316; PubMed=1537604;
RA   Norimine J., Miyazawa T., Kawaguchi Y., Tohya Y., Kai C., Mikami T.;
RT   "A cDNA encoding feline CD4 has a unique repeat sequence downstream of
RT   the V-like region.";
RL   Immunology 75:74-79(1992).
DR   EMBL; AB000483; BAA19124.1; -.
DR   HSSP; P01730; 1WIQ.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 3.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 2.
KW   Signal.
FT   SIGNAL       1     26       POTENTIAL.
FT   CHAIN       27    474       POTENTIAL.
SQ   SEQUENCE   474 AA;  52243 MW;  D946DD4BEAED00EC CRC64;

  Query Match             46.8%;  Score 1087;  DB 6;  Length 474;
  Best Local Similarity   57.6%;  Pred. No. 4.2e-75;
  Matches  239;  Conservative   49;  Mismatches   99;  Indels   28;  Gaps    6;

Qy          1 MNRGVPFRHLLLVLQLALLPAAT-QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQI 59
              ||:|  |||||||||| :| ||  || :||||| | | || | ||||| : | |: |:|:
Db          1 MNQGAVFRHLLLVLQLVMLKAAVPQGKEVVLGKAGGTAELPCQASQKKYMTFTWRLSSQV 60

Qy         60 KILGNQGSFL-TKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEV 118
              ||| :| | |   | |||  | :|:: |||||:|||:||:|:: ||  | ||||::| ||
Db         61 KILESQHSSLCLTGSSKLKTRFESKKILWDQGSFPLVIKSLQVADSGIYTCEVENKKREV 120

Qy        119 QLLVFGLTANSD----------------THLLQGQSLTLTLESPPGSSPSVQCRSPRGK 161
              :|||||||  |                :|||||||||||||  |:|||| : |  |
Db        121 ELLVFGLTAKVDPSGSGGSSSSSTSTSTSIYLLQGQSLTLTLESPSSSNPSVQWKGPGNK 180

Qy        162 NIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEF 221
              :  |  :||:||||||:||| |||| |:|| : |  :|:||||:| |: || ||||||||
Db        181 SKSGVHSLSLSQLELQESGTCTCTVSQSQKTLVFNTNILVLAFRKVSNTVYAKEGEQVEF 240

Qy        222 SFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLH 281
              |||| |  | | |:  | |:|| | ||  ||:| ||||::||| |    ||||   |||
Db        241 SFPLNFEDENLMGN--LRWKAEGAPSSLLWISFTLKNKQLSVKEVDPYSKLQMMDSLPLR 298

Qy        282 LTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLS 341
              |||  | :|||||||| |:   |:| ||| ||||| ||   |||||| |||||:| ||
Db        299 FTLPNVLSRYAGSGNLTLVLD--KGQLQQEVKLVVMRVTQSGNNLTCEVLGPTSPELTLS 356

Qy        342 LKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVLP-----TWS 391
              |||: : |||||::| | | : ||| ||||||   :||| |  :|||     ||:
Db        357 LKLKGQAAKVSKQQKMVRVEDAEAGTWQCLLSHKDKVLLASKAEVLPPVLTRTWT 411


RESULT 12
Q61396
ID   Q61396      PRELIMINARY;      PRT;   457 AA.
AC   Q61396;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell differentiation antigen.
GN   CD4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88041159; PubMed=2823269;
RA   Gorman S.D., Tourvieille B., Parnes J.R.;
RT   "Structure of the mouse gene encoding CD4 and an unusual transcript in
RT   brain.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).
DR   EMBL; M17080; AAA37402.1; -.
DR   EMBL; M17076; AAA37402.1; JOINED.
DR   EMBL; M17077; AAA37402.1; JOINED.
DR   EMBL; M17078; AAA37402.1; JOINED.
DR   EMBL; M17079; AAA37402.1; JOINED.
DR   HSSP; P01730; 1WBR.
DR   MGD; MGI:88335; Cd4.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
SQ   SEQUENCE   457 AA;  51368 MW;  24AB19EDA285B5D0 CRC64;

  Query Match           42.4%;  Score 985;  DB 11;  Length 457;
  Best Local Similarity   53.6%;  Pred. No. 2.7e-67;
  Matches  210;  Conservative   62;  Mismatches  110;  Indels   10;  Gaps    7;

Qy          1 MNRGVPFRH-LLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQI 59
              | | :  |  |||:|||: | | ||   :||||:|:: || | :||||    | || |:|
Db          1 MCRAISLRRLLLLLLQLSQLLAVTQEKTLVLGKEGESAELPCESSQKKITVFTWKFSDQR 60

Qy         60 KILGNQG-SFLTKG--PSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKE 116
              ||||  |    | :|  ||:  || ||::  |::|:|||||  ||:||| |||||:|::||
Db         61 KILGQHGKGVLIRGGSPSQF-DRFDSKKGAWEKGSFPLIINKLKMEDSQTYICELENRKE 119

Qy        117 EVQLLVFGLTANSDTHLLQGQSLTLTLES-PPGSSPSVQCRSPRGKNIQGGKTLSVSQLE 175
              ||:| || :| :  | ||||||||||||:|    |:|  :|:  :|| : | | ||:| |
Db        120 EVELWVFKVTFSPGTSLLQGQSLTLTLDSNSKVSNPLTECKHKKGKVVSGSKVLSMSNLR 179

Qy        176 LQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGS 235
              :|||  | |||  :|||  | : : || ||  :   || |||  ||||||  |  |   |
Db        180 VQDSDFWNCTVTLDQKKNWFGMTLSVLGFQSTAITAYKSEGESAEFSFPLNFAEE--NGW 237

Qy        236 GELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSG 295
              ||| |:||: |  : ||:| :|||||||:: |:|  |||: : ||| | :||   |:||||
Db        238 GELMWKAEKDSFFQPWISFSIKNKEVSVQKSTKDLKLQLKETLPLTLKIPQVSLQFAGSG 297

Qy        296 NLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKRE 355
              |||| |:   | ||||||||||:  ||   |||| |||||: |:|| ||:||:||: :
Db        298 NLTLTLD--KGTLHQEVNLVVMKVAQLNNTLTCEVMGPTSPKMRLTLKQENQEARVSEEQ 355

Qy        356 KPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
              | | |: || |:||||||:  :| ::| |:||
Db        356 KVVQVVAPETGLWQCLLSEGDKVKMDSRIQVL 387


RESULT 13
O55054
ID   O55054      PRELIMINARY;      PRT;   433 AA.
AC   O55054;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T4 surface glycoprotein (Fragment).
GN   CD4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88097446; PubMed=3501122;
RA   Maddon P.J., Molineaux S.M., Maddon D.E., Zimmerman K.A., Godfrey M.,
RA   Alt F.W., Chess L., Axel R.;
RT   "Structure and expression of the human and mouse T4 genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:9155-9159(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Maddon P.J., Molineaux S.M., Maddon D.E., Zimmerman K.A., Godfrey M.,
RA   Alt F.W., Chess L., Axel R.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF045882; AAC01764.1; -.
DR   HSSP; P01730; 1WBR.
DR   MGD; MGI:88335; Cd4.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 2.
DR   PRINTS; PR00692; CD4TCANTIGEN.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER       1      1
FT   NON_TER     433    433
SQ   SEQUENCE   433 AA;  48590 MW;  AB19330750A8499A CRC64;

  Query Match           41.1%;  Score 954.5;  DB 11;  Length 433;
  Best Local Similarity   54.1%;  Pred. No. 5.6e-65;
  Matches  199;  Conservative   59;  Mismatches  101;  Indels    9;  Gaps    6;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQG-SFLTKG--PSKLNDRA 80
              ||  :||||:|:: || | :||||    | || |:| ||||  |    | :|  ||:  ||
Db          1 QGKTLVLGKEGESAELPCESSQKKITVFTWKFSDQRKILGQHGKGVLIRGGSPSQF-DRF 59

Qy         81 DSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLT 140
              ||::  |::|:|||||  ||:||| |||||:|::||||:| || :| :  | ||||||||
Db         60 DSKKGAWEKGSFPLIINKLKMEDSQTYICELENRKEEVELWVFKVTFSPGTSLLQGQSLT 119

Qy        141 LTLES-PPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDI 199
              |||:|    |:|  :|:  :|| : | | ||:| | :|||  | |||  :|||  | : :
Db        120 LTLDSNSKVSNPLTECKHKKGKVVSGSKVLSMSNLRVQDSDFWNCTVTLDQKKNWFGMTL 179

Qy        200 VVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNK 259
              || ||  :   || |||  |||||| |  |   | ||| |:||: |  : ||:| :|||
Db        180 SVLGFQSTAITAYKSEGESAEFSFPLNFAEE--NGWGELMWKAEKDSFFQPWISFSIKNK 237

Qy        260 EVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRA 319
              ||||:: |:| ||: : ||| | :||   |:||||||||| |:   | ||||||||||:
Db        238 EVSVQKSTKDLKLQLKETLPLTLKIPQVSLQFAGSGNLTLTLD--KGTLHQEVNLVVMKV 295

Qy        320 TQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVL 379
              ||   ||||| |||||: |:|| ||:||:||: :| | |: || |:||||||:  :|
Db        296 AQLNNTLTCEVMGPTSPKMRLTLKQENQEARVSEEQKVVQVVAPETGLWQCLLSEGDKVK 355

Qy        380 LESNIKVL 387
              ::| |:||
Db        356 MDSRIQVL 363


RESULT 14
O77596
ID   O77596      PRELIMINARY;      PRT;    86 AA.
AC   O77596;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE   (Fragment).
OS   Mandrillus sphinx (Mandrill) (Papio sphinx).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Mandrillus.
OX   NCBI_TaxID=9561;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98320644; PubMed=9656488;
RA   Harris E.E., Disotell T.R.;
RT   "Nuclear gene trees and the phylogenetic relationships of the
RT   mangabeys (primates: Papionini).";
RL   Mol. Biol. Evol. 15:892-900(1998).
CC   -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC       RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR   EMBL; AF057386; AAC25130.1; -.
DR   HSSP; P01730; 1CDY.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   InterPro; IPR007110; Ig-like.
KW   Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT   NON_TER       1      1
FT   DOMAIN       <1     19       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       20     >86      IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID     49     78       BY SIMILARITY.
FT   NON_TER      86     86
SQ   SEQUENCE   86 AA;  9406 MW;  2BD97A9EE19582AB CRC64;

  Query Match             17.2%;  Score 400;  DB 6;  Length 86;
  Best Local Similarity   89.4%;  Pred. No. 2.3e-23;
  Matches   76;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
              |||||||:||||:|||||||||||||||:||||||||||||||||:||||||||||||
Db          1 YICEVEDKKEEVELLVFGLTANSDTHLLEGQSLTLTLESPPGSSPSVKCRSPRGKNIQGG 60

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQK 191
              :|||| ||| |||||||||| |:||
Db         61 RTLSVPQLERQDSGTWTCTVSQDQK 85


RESULT 15
O77597
ID   O77597      PRELIMINARY;      PRT;    86 AA.
AC   O77597;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE   (Fragment).
OS   Mandrillus leucophaeus (Drill) (Papio leucophaeus).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Mandrillus.
OX   NCBI_TaxID=9568;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98320644; PubMed=9656488;
RA   Harris E.E., Disotell T.R.;
RT   "Nuclear gene trees and the phylogenetic relationships of the
RT   mangabeys (primates: Papionini).";
RL   Mol. Biol. Evol. 15:892-900(1998).
CC   -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC       RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR   EMBL; AF057387; AAC25131.1; -.
DR   HSSP; P01730; 1CDY.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   InterPro; IPR007110; Ig-like.
KW   Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT   NON_TER       1      1
FT   DOMAIN       <1     19       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       20     >86      IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID     49     78       BY SIMILARITY.
FT   NON_TER      86     86
SQ   SEQUENCE   86 AA;  9406 MW;  2BD97A9EE19582AB CRC64;

  Query Match             17.2%;  Score 400;  DB 6;  Length 86;
  Best Local Similarity   89.4%;  Pred. No. 2.3e-23;
  Matches   76;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
              |||||||:||||:|||||||||||||||:||||||||||||||||:||||||||||||
Db          1 YICEVEDKKEEVELLVFGLTANSDTHLLEGQSLTLTLESPPGSSPSVKCRSPRGKNIQGG 60

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQK 191
              :|||| ||| |||||||||| |:||
Db         61 RTLSVPQLERQDSGTWTCTVSQDQK 85


RESULT 16
O77594
ID   O77594      PRELIMINARY;      PRT;    86 AA.
AC   O77594;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE   (Fragment).
OS   Cercopithecus mitis.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=36225;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98320644; PubMed=9656488;
RA   Harris E.E., Disotell T.R.;
RT   "Nuclear gene trees and the phylogenetic relationships of the
RT   mangabeys (primates: Papionini).";
RL   Mol. Biol. Evol. 15:892-900(1998).
CC   -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC       RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR   EMBL; AF057381; AAC25125.1; -.
DR   HSSP; P01730; 1CDY.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   InterPro; IPR007110; Ig-like.
KW   Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT   NON_TER       1      1
FT   DOMAIN       <1     19       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       20     >86      IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID     49     78       BY SIMILARITY.
FT   NON_TER      86     86
SQ   SEQUENCE   86 AA;  9417 MW;  371CA39EF58182AB CRC64;

  Query Match             17.1%;  Score 397;  DB 6;  Length 86;
  Best Local Similarity   89.4%;  Pred. No. 3.9e-23;
  Matches   76;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
              |||||||:||||:|||||||||||||||:||||||||||||||||:||||||||||||
Db          1 YICEVEDKKEEVELLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVKCRSPRGKNIQGG 60

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQK 191
              : ||| ||| |||||||||| |:||
Db         61 RILSVPQLERQDSGTWTCTVSQDQK 85


RESULT 17
```

```
O77599
ID   O77599      PRELIMINARY;      PRT;    86 AA.
AC   O77599;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE   (Fragment).
OS   Theropithecus gelada (Gelada baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Theropithecus.
OX   NCBI_TaxID=9565;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98320644; PubMed=9656488;
RA   Harris E.E., Disotell T.R.;
RT   "Nuclear gene trees and the phylogenetic relationships of the
RT   mangabeys (primates: Papionini).";
RL   Mol. Biol. Evol. 15:892-900(1998).
CC   -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC       RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR   EMBL; AF057389; AAC25133.1; -.
DR   HSSP; P01730; 1CDY.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   InterPro; IPR007110; Ig-like.
KW   Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT   NON_TER       1      1
FT   DOMAIN       <1     19       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       20    >86       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID     49     78       BY SIMILARITY.
FT   NON_TER      86     86
SQ   SEQUENCE   86 AA;  9420 MW;  2BD97A98574EE9AB CRC64;

  Query Match             17.1%;  Score 397;  DB 6;  Length 86;
  Best Local Similarity   88.2%;  Pred. No. 3.9e-23;
  Matches   75;  Conservative    7;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
              |||||||:||||:||||||||||||||||:|||||||||||||:||||:||||||||||||
Db          1 YICEVEDKKEEVELLVFGLTANSDTHLLEGQSLTLTLESPPGTSPSVKCRSPRGKNIQGG 60

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQK 191
              :|||| ||| |||||||||| |:||
Db         61 RTLSVPQLERQDSGTWTCTVSQDQK 85


RESULT 18
O77595
ID   O77595      PRELIMINARY;      PRT;    86 AA.
AC   O77595;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE   (Fragment).
OS   Cercocebus galeritus chrysogaster.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercocebus.
OX   NCBI_TaxID=75569;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98320644; PubMed=9656488;
RA   Harris E.E., Disotell T.R.;
RT   "Nuclear gene trees and the phylogenetic relationships of the
RT   mangabeys (primates: Papionini).";
RL   Mol. Biol. Evol. 15:892-900(1998).
CC   -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC       RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR   EMBL; AF057382; AAC25126.1; -.
DR   HSSP; P01730; 1CDY.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   InterPro; IPR007110; Ig-like.
KW   Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT   NON_TER       1      1
FT   DOMAIN       <1     19       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       20    >86       IG-LIKE C2-TYPE DOMAIN.
FT   CARBOHYD     79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     49     78       BY SIMILARITY.
FT   NON_TER      86     86
SQ   SEQUENCE   86 AA;  9419 MW;  A9D97A9EE19582BE CRC64;

  Query Match             17.0%;  Score 395;  DB 6;  Length 86;
  Best Local Similarity   88.2%;  Pred. No. 5.6e-23;
  Matches   75;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
              |||||||:||||:||||||||||||||||:|||||||||||||:||||:||||||||||||
Db          1 YICEVEDKKEEVELLVFGLTANSDTHLLEGQSLTLTLESPPGSSPSVKCRSPRGKNIQGG 60

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQK 191
              :|||| ||| |||||||| | |:||
Db         61 RTLSVPQLERQDSGTWTCNVSQDQK 85


RESULT 19
O77598
ID   O77598      PRELIMINARY;      PRT;    86 AA.
AC   O77598;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE   (Fragment).
OS   Papio sp. (Baboon).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=61183;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98320644; PubMed=9656488;
RA   Harris E.E., Disotell T.R.;
RT   "Nuclear gene trees and the phylogenetic relationships of the
RT   mangabeys (primates: Papionini).";
RL   Mol. Biol. Evol. 15:892-900(1998).
CC   -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC       RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR   EMBL; AF057388; AAC25132.1; -.
DR   HSSP; P01730; 1CDY.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   InterPro; IPR007110; Ig-like.
KW   Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT   NON_TER       1      1
FT   DOMAIN       <1     19       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       20    >86       IG-LIKE C2-TYPE DOMAIN.
FT   CARBOHYD     79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     49     78       BY SIMILARITY.
FT   NON_TER      86     86
SQ   SEQUENCE   86 AA;  9433 MW;  A9D97A98574EE9BE CRC64;
```

```
  Query Match             16.9%;  Score 392;  DB 6;  Length 86;
  Best Local Similarity   87.1%;  Pred. No. 9.5e-23;
  Matches   74;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
              |||||||:||||:||||||||||||||||:|||||||||||||:||||:||||||||||||
Db          1 YICEVEDKKEEVELLVFGLTANSDTHLLEGQSLTLTLESPPGTSPSVKCRSPRGKNIQGG 60

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQK 191
              :|||| ||| |||||||| | |:||
Db         61 RTLSVPQLERQDSGTWTCNVSQDQK 85


RESULT 20
O77601
ID   O77601      PRELIMINARY;      PRT;    86 AA.
AC   O77601;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE   (Fragment).
OS   Lophocebus albigena albigena.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Lophocebus.
OX   NCBI_TaxID=75568;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98320644; PubMed=9656488;
RA   Harris E.E., Disotell T.R.;
RT   "Nuclear gene trees and the phylogenetic relationships of the
RT   mangabeys (primates: Papionini).";
RL   Mol. Biol. Evol. 15:892-900(1998).
CC   -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC       RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR   EMBL; AF057391; AAC25135.1; -.
DR   HSSP; P01730; 1CDY.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   InterPro; IPR007110; Ig-like.
KW   Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT   NON_TER       1      1
FT   DOMAIN       <1     19       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       20    >86       IG-LIKE C2-TYPE DOMAIN.
FT   DISULFID     49     78       BY SIMILARITY.
FT   NON_TER      86     86
SQ   SEQUENCE   86 AA;  9463 MW;  2BD97A88464FE9AB CRC64;

  Query Match             16.7%;  Score 388;  DB 6;  Length 86;
  Best Local Similarity   87.1%;  Pred. No. 1.9e-22;
  Matches   74;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
              |||||||:||||:||||||||||||||||:|||||||||||||:||||:|||||||||| |
Db          1 YICEVEDKKEEVELLVFGLTANSDTHLLEGQSLTLTLESPPGTSPSVKCRSPRGKNIQVG 60

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQK 191
              :|||| ||| |||||||||| |:||
Db         61 RTLSVPQLERQDSGTWTCTVSQDQK 85


RESULT 21
O77600
ID   O77600      PRELIMINARY;      PRT;    86 AA.
AC   O77600;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE   (Fragment).
OS   Lophocebus aterrimus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Lophocebus.
OX   NCBI_TaxID=75566;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98320644; PubMed=9656488;
RA   Harris E.E., Disotell T.R.;
RT   "Nuclear gene trees and the phylogenetic relationships of the
RT   mangabeys (primates: Papionini).";
RL   Mol. Biol. Evol. 15:892-900(1998).
CC   -!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC       RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC       ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR   EMBL; AF057390; AAC25134.1; -.
DR   HSSP; P01730; 1CDY.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   InterPro; IPR007110; Ig-like.
KW   Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT   NON_TER       1      1
FT   DOMAIN       <1     19       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       20    >86       IG-LIKE C2-TYPE DOMAIN.
FT   CARBOHYD     79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     49     78       BY SIMILARITY.
FT   NON_TER      86     86
SQ   SEQUENCE   86 AA;  9476 MW;  A9D97A88464FE9BE CRC64;

  Query Match             16.5%;  Score 383;  DB 6;  Length 86;
  Best Local Similarity   85.9%;  Pred. No. 4.7e-22;
  Matches   73;  Conservative    7;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
              |||||||:||||:||||||||||||||||:|||||||||||||:||||:|||||||||| |
Db          1 YICEVEDKKEEVELLVFGLTANSDTHLLEGQSLTLTLESPPGTSPSVKCRSPRGKNIQVG 60

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQK 191
              :|||| ||| |||||||| | |:||
Db         61 RTLSVPQLERQDSGTWTCNVSQDQK 85


RESULT 22
Q13969
ID   Q13969      PRELIMINARY;      PRT;    71 AA.
AC   Q13969;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4 protein (Fragment).
GN   CD4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zverev V.V., Blinov V.M., Nedospasov S.A.;
RT   "Splice-mediated insertion of antisence and sence Alu repeats in human
RT   CD4 gene: identification of three exons of CD4 mRNA.";
RL   Submitted (APR-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95407135; PubMed=7676667;
RA   Zverev V.V., Sidorov A.V., Nedospasov S.A., Maliushova V.V.,
RA   Udalova I.A., Andzhaparizde O.G., Blinov V.M.;
RT   "[Nucleotide sequence of two exons of the human T-lymphocyte CD4
```

```
RT   receptor gene].";
RL   Vopr. Virusol. 40:100-102(1995).
DR   EMBL; X87579; CAA60883.1; -.
DR   EMBL; S79267; AAB35273.1; -.
DR   PIR; I60082; I60082.
DR   HSSP; P01730; 1CDY.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   InterPro; IPR007110; Ig-like.
DR   PROSITE; PS50835; IG_LIKE; 1.
KW   Receptor.
FT   NON_TER      71     71
SQ   SEQUENCE   71 AA;  7844 MW;  A5C9D84816135C86 CRC64;

  Query Match             15.4%;  Score 357;  DB 4;  Length 71;
  Best Local Similarity   98.6%;  Pred. No. 3.6e-20;
  Matches   70;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTK 71
              ||||||||||||
Db         61 ILGNQGSFLTK 71


RESULT 23
Q29027
ID   Q29027      PRELIMINARY;      PRT;    99 AA.
AC   Q29027;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4, allele 1 (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=#1183; TISSUE=Blood;
RX   MEDLINE=93329116; PubMed=8335933;
RA   Gustafsson K., Germana S., Sundt T.M., Sachs D.H., LeGuern C.;
RT   "Extensive allelic polymorphion in an exposed region of the
RT   miniature. . . . . .";
RL   J. Immunol. 151:1365-1370(1993).
DR   EMBL; X65629; CAA46583.1; -.
DR   PIR; I47131; S21461.
DR   HSSP; P01730; 1CDY.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER       1      1
FT   NON_TER      99     99
SQ   SEQUENCE   99 AA;  11170 MW;  40BF080699CF5D0C CRC64;

  Query Match             13.1%;  Score 305.5;  DB 6;  Length 99;
  Best Local Similarity   60.2%;  Pred. No. 5.1e-16;
  Matches   59;  Conservative   16;  Mismatches   22;  Indels    1;  Gaps    1;

Qy         32 KKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFL-TKGPSKLNDRADSRRSLWDQG 90
              | ||  || | :||||:: |:|||||| ||||  |||  |   ::|  | ||::::|| |
Db          1 KAGDLAELPCHSSQKKNLPFNWKNSNQTKILGGHGSFWHTASVTELTSRLDSKKNMWDHG 60

Qy         91 NFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTAN 128
              :||||||||:: ||  |||||||:: ||||||| |||:
Db         61 SFPLIIKNLEVTDSGIYICEVEDKRIEVQLLVFRLTAS 98


RESULT 24
Q29028
ID   Q29028      PRELIMINARY;      PRT;    99 AA.
AC   Q29028;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4, allele 2 (Fragment).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=#1183; TISSUE=Blood;
RX   MEDLINE=93329116; PubMed=8335933;
RA   Gustafsson K., Germana S., Sundt T.M., Sachs D.H., LeGuern C.;
RT   "Extensive allelic polymorphion in an exposed region of the
RT   miniature. . . . . .";
RL   J. Immunol. 151:1365-1370(1993).
DR   EMBL; X65630; CAA46584.1; -.
DR   PIR; I47132; S21462.
DR   HSSP; P01730; 1CDY.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
FT   NON_TER       1      1
FT   NON_TER      99     99
SQ   SEQUENCE   99 AA;  11390 MW;  C229E5BEA228318F CRC64;

  Query Match             12.1%;  Score 280.5;  DB 6;  Length 99;
  Best Local Similarity   56.1%;  Pred. No. 4.3e-14;
  Matches   55;  Conservative   19;  Mismatches   23;  Indels    1;  Gaps    1;

Qy         32 KKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGP-SKLNDRADSRRSLWDQG 90
              | ||  || | :||||:: | ||||:||||| :  :   |   ::|: | ||::::|| |
Db          1 KAGDLAELPCHSSQKKNLPFSWKNSDQIKILRSHRNLWHKASVTELSSRLDSKKNMWDHG 60

Qy         91 NFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTAN 128
              :||||||||:: ||  |||||||:: ||||||| |||:
Db         61 SFPLIIKNLEVTDSGIYICEVEDKRIEVQLLVFRLTAS 98


RESULT 25
O88650
ID   O88650      PRELIMINARY;      PRT;   120 AA.
AC   O88650;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 (Fragment).
GN   CD4.
OS   Marmota monax (Woodchuck).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC   Marmota.
OX   NCBI_TaxID=9995;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Peripheral blood;
RX   MEDLINE=20094937; PubMed=10627561;
RA   Guo J.T., Zhou H., Liu C., Aldrich C., Saputelli J., Whitaker T.,
RA   Barrasa M.I., Mason W.S., Seeger C.;
RT   "Apoptosis and regeneration of hepatocytes during recovery from
RT   transient hepadnavirus infections.";
RL   J. Virol. 74:1495-1505(2000).
DR   EMBL; AF082497; AAC32621.1; -.
DR   HSSP; P01730; 1WIO.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0006955; P:immune response; IEA.
DR   InterPro; IPR000973; CD4_TCAg.
DR   InterPro; IPR007110; Ig-like.
DR   PRINTS; PR00692; CD4TCANTIGEN.
```

```
FT   NON_TER       1      1
FT   NON_TER     120    120
SQ   SEQUENCE   120 AA;  13623 MW;  44967B512D520195 CRC64;

  Query Match             12.0%;  Score 278.5;  DB 11;  Length 120;
  Best Local Similarity   48.3%;  Pred. No. 8e-14;
  Matches   71;  Conservative    9;  Mismatches   38;  Indels   29;  Gaps    3;

Qy        308 LHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGM 367
              ||||||||||  |  | :| ||| ||| |||||||||:|:|||||| || : | ||:|||
Db          1 LHQEVNLVVMNVTYHQNDLICEVLGPTPPKLMLSLKLKNQEAKVSKPEKRIRVPNPKAGM 60

Qy        368 WQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASA 427
              |||||  |  :|||:  | | ||       |    |:     | ||
Db         61 WQCLLRDGDKVLLDFQIDVPPTELNQYQPMFLAV---IIGGAL---------------- 100

Qy        428 LPAALAVISFLLGLGLGV-ACVLARTR 453
                      ||||  || :  ||  | |
Db        101 --------SFLLLAGLCIFCCVKCRHR 119


RESULT 26
Q90WB5
ID   Q90WB5      PRELIMINARY;      PRT;   482 AA.
AC   Q90WB5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   T-cell surface glycoprotein CD4 precursor.
OS   Anas platyrhynchos (Domestic duck).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX   NCBI_TaxID=8839;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RA   Chan S.W.S., Middleton D.L., Lundqvist M., Warr G.W., Higgins D.A.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF378701; AAK59279.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 4.
DR   SMART; SM00409; IG; 3.
DR   PROSITE; PS50835; IG_LIKE; 2.
KW   Signal.
FT   SIGNAL        1     29       POTENTIAL.
SQ   SEQUENCE   482 AA;  54624 MW;  BDA0E8EC6172AD0B CRC64;

  Query Match             11.7%;  Score 271.5;  DB 13;  Length 482;
  Best Local Similarity   25.0%;  Pred. No. 1.9e-12;
  Matches  111;  Conservative   69;  Mismatches  169;  Indels   95;  Gaps   19;

Qy          8 RHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWK-------NSNQIK 60
              | : ::||| |       ::  :| :|  | | |   :|          ::   |:
Db         12 RAVFVLLQLGL--THIMAHQQQIGVEGKEVILNC---KKHDKDVTWKYEYDAGSSAIIIQ 66

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||  :   : || : ::||:::     :| :  | : ||:| |: |||||    :  : |
Db         67 ILAGK---IFKGRAPMSDRSET-----NQNSKHLKVSNLRISDAGTYICECGSDRNSISL 118

Qy        121 LVFGLTANSDTHLLQGQSLTLT-LESPPGSSPS-----------------VQCRSPRGKN 162
               |  || :|: : | |  | || :   | | |                  :|  :|:
Db        119 HVVKLTISSNGYFLPGDDLELTVMHKSPKSQPRFSITLFNSHNSRVTPEVLQNETPQ--- 175

Qy        163 IQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKI--DIVVLAFQKAS-SIVYKKEGEQV 219
              | | ||:  |||| | :   :    :   :: || |:|     :|        |
Db        176 ---KYALKVKQLQPTDSGTWICNMHSDSPSINENISFNVKVLGFEKTHLERMYAAVDSTV 232

Qy        220 EFSFPLAFTVEKLTG-----SGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQM 274
              |: | |   :  |     :|:| ||   |      ||::| :   :     :: |
Db        233 TLSWHLNF---RKIGWKEFFTGQLNWQEGNA------ITYELLDFNATADGELRETK--- 280

Qy        275 GKKLPLHLTLPQALP----------QYAGSGNLTLALEAKTGKLHQEVNLVVMRATQ-- 321
                |:    | :|:              |   ||  |  |      :  :  ||||: :
Db        281 -KRSQALLEIPEMKRDSTVEVKIHKIQLKHSGEYTCQLLYNRRYIQSKTELVVMQVSANP 339

Qy        322 -------LQKNLTCEVWGPTSPKLML------SLKLENKEAKVSKREKPVWVLNPEAGMW 368
                    :  | |:|  |  | : |        |:: |::| |  |  | | |     |||
Db        340 PGPLPKGAEMTLLCQVSSPIPPNVHLLWERVNGTKMDGKKSKQS--ETKVEVKVTAVGMW 397

Qy        369 QCLLSDSGQVLLESN--IKVLPTW 390
                | | :   : | |  ::  |||
Db        398 NCHLMEDNNMKLSLNYTVEEAPTW 421


RESULT 27
Q9W6V7
ID   Q9W6V7      PRELIMINARY;      PRT;   487 AA.
AC   Q9W6V7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4 protein precursor.
GN   CD4.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H.B2; TISSUE=Thymus;
RX   MEDLINE=99218434; PubMed=10201936;
RA   Koskinen R., Lamminmaki U., Tregaskes C.A., Salomonsen J., Young J.R.,
RA   Vainio O.;
RT   "Cloning and modeling of the first nonmammalian CD4.";
RL   J. Immunol. 162:4115-4121(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B12;
RA   Koskinen R.H., Salomonsen J., Tregaskes C.A., Young J.R.,
RA   Goodchild M., Bumstead N., Vainio O.;
RT   "The chicken CD4 gene has remained conserved in evolution.";
RL   Immunogenetics 0:0-0(2002).
DR   EMBL; Y12012; CAA72740.1; -.
DR   EMBL; AJ401223; CAC82027.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 3.
DR   PROSITE; PS50835; IG_LIKE; 2.
KW   Signal.
FT   SIGNAL        1     28       POTENTIAL.
SQ   SEQUENCE   487 AA;  54985 MW;  B9CBA92EC9F7F45B CRC64;

  Query Match             10.5%;  Score 245;  DB 13;  Length 487;
  Best Local Similarity   23.9%;  Pred. No. 2.1e-10;
  Matches  121;  Conservative   66;  Mismatches  193;  Indels  126;  Gaps   22;

Qy         10 LLLVLQLALLPAATQGNKVVLGKKGDTVELTCTA-SQKKSIQFHWKN-----SNQIKILG 63
              ::|||| | |    |  : : | |  | |:| | : :|     ||      |: |
Db         14 VILVLQLGLTPIMAQQEQQI-GIAGKEVILSCKAINNQKDGTCTWKYKYKEVSSTIISFS 72

Qy         64 NQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVF 123
                 | | |          : ||  :  :  | : :| ::|:  | |         : | ||
Db         73 KAQVFKGKAPM-------THRSELNSNSKKLKVSDLSLDDAGIYTCACYSPVVSISLHVF 125

Qy        124 GLTANSDTHLLQGQSLTLTL-ESPPGSSPS-----------------VQCRSPRGKNIQG 165
               || :|: | |  : | |||  ::   | |                  :|  :|: | |
Db        126 KLTISSNGHFLTNEDLELTLMQNSSHSQPHLSIKLFNINNDIVTTEILQEEAPQ-KYI-- 182

Qy        166 GKTLSVSQLELQDSGTWTCTVLQNQKKVEFKI--DIVVLAFQKAS-SIVYKKEGEQVEFS 222
              | : ||:  |||| | |  |    :   |  |: || |:|    |:|   |      |
Db        183 ---LKLKQLKAIDSGTWMCHVYSNSPSINQNISFDVKVLGFEKERLEIIYTTVGNTAILS 239

Qy        223 FPLAFTVEKLTG--SGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPL 280
              : | |   |      :|:| |: :  ::     : |              ||   :|   ||
Db        240 WRLNFRKIKWKEGFTGKLNWEPQGNTAIHELLNFS----------VTTHQELHKTKK-SN 288
```

```
Qy        281 HL-----------TLPQALP--QYAGSGNLTLALEAKTGKLHQEVNLVVMRATQL----- 322
              |:           |:   :|  |   ||     ||    :     ||||: | :
Db        289 HIWFEISEGKTDGTMDVKIPKVQLNHSGQYKCQLEINGRRTESVRALVVMQVTAIPAGPL 348

Qy        323 ----QKNLTCEVWGPTSPKLMLSLKLEN---KEAKVSKREKPVWVLNPEA-GMWQC-LLS 373
                  :  | |:| ||      |  :  |    | | ||: :    :|  | |:| | |:
Db        349 SRGGKMTLLCQVSGPLPSNAHLLWERVNGTQMEMKKSKQHEAKVEVNVSAPGLWNCHLVE 408

Qy        374 DSG-QVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAAL 432
              |:  :: |   ::    |:                                       :
Db        409 DNNKKISLNYTVEEAHVWN----------------------------------------SY 429

Qy        433 AVISFLLG-----LGLGVACVLARTR 453
              |||  ::|     :||   |::   |
Db        430 AVIGIIIGASVLVIGLACMCIITGMR 455


RESULT 28
P70443
ID   P70443      PRELIMINARY;      PRT;    90 AA.
AC   P70443;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mutant T-cell surface glycoprotein CD4 (Fragment).
GN   CD4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BR/cdJ;
RA   Rosenwasser O.A., Fairchild S.P., Tomonari K.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U75219; AAB18133.1; -.
DR   HSSP; P05540; 1CID.
DR   InterPro; IPR007110; Ig-like.
FT   NON_TER      1      1
SQ   SEQUENCE   90 AA;  9938 MW;  6528B707928A4685 CRC64;

  Query Match             10.2%;  Score 237;  DB 11;  Length 90;
  Best Local Similarity   59.2%;  Pred. No. 8.2e-11;
  Matches   45;  Conservative   14;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        310 QEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQ 369
              ||||||||:  ||   |||||  ||||||: |:|| ||:||:||: :| | |: || |:||
Db          1 QEVNLVVMKVAQLNNTLTCEVMGPTSPKMRLTLKQENQEARVSEEQKVVQVVAPETGLWQ 60

Qy        370 CLLSDSGQVLLESNIK 385
              ||||:  :| ::| |:
Db         61 CLLSEGDKVKMDSRIQ 76


RESULT 29
Q28260
ID   Q28260      PRELIMINARY;      PRT;   739 AA.
AC   Q28260;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vascular cell adhesion molecule-1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ballantyne C.M., Clubb F.J., Perrard J.L., Radovencovic B.,
RA   Youker K.A., Smith C.W., Entman M.L., Hawkins H.K., Frazier O.H.,
RA   Willerson J.T.;
RT   "Increased Expression of VCAM-1 and ICAM-1 in Early Cardiac Allograft
RT   Arteriopathy in the Dog.";
RL   Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U32086; AAA84866.1; -.
DR   HSSP; P19320; 1VCA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016337; P:cell-cell adhesion; IEA.
DR   InterPro; IPR003987; ICAM_VCAM-1.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003989; VCAM-1.
DR   Pfam; PF00047; ig; 7.
DR   PRINTS; PR01472; ICAMVCAM1.
DR   PRINTS; PR01474; VCAM1.
DR   SMART; SM00408; IGc2; 4.
DR   PROSITE; PS50835; IG_LIKE; 5.
KW   Immunoglobulin domain.
SQ   SEQUENCE   739 AA;  81412 MW;  BB5DA3853739C615 CRC64;

  Query Match             7.5%;  Score 174;  DB 6;  Length 739;
  Best Local Similarity   22.5%;  Pred. No. 0.00011;
  Matches   86;  Conservative   74;  Mismatches  140;  Indels   82;  Gaps   19;

Qy         32 KKGDTVELTCTASQKKSIQFHWK---NSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWD 88
              ::| :| :|| :       |  |    ::  :::|
Db        237 QEGGSVTMTCASEGLPPPQIFWSKKLDNGNLQLL-------------------------- 270

Qy         89 QGNFPLIIKNLKIEDSDTYIC----EVEDQKEEVQLLV----FGLTANSDTHLLQ--GQS 138
               ||  | :  :::||| ||:|    ||    :||:|:|    | :  :   ::   | |
Db        271 SGNATLTLIAMRLEDSGTYVCEGVNEVGKDGKEVELIVQEKPFTVEISPGPQIIAQIGDS 330

Qy        139 LTLTLESPPGSSPS----VQCRSPRG--KNIQGGK-TLSVSQLELQDSGTWTCTVLQNQK 191
              : ||        |||      |  ||       ::| | ||::| : |::  :: |||    |
Db        331 VVLTCGVTDCESPSFSWRTQIDSPLSGTVKVEGAKSTLTLSPVNLENEHSYLCTVTCGHK 390

Qy        192 KVEFKIDIVVLAFQK-----ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERAS 246
              |:|  | : : :| :       | ::        |     |  :  ::|    || :: |
Db        391 KLEKGIKVDLYSFPRDPEVEMSGLLVDGNPVTVSCEVPNVYPSDRL--EIEL-FKGETII 447

Qy        247 SSKSWI------TFDLKNKEVSVKRVTQDPKLQMGKKL----PLHLTLPQALPQYAGSGN 296
              |||::      : : |: }::    |:|     || |     ||:   :   |:   |
Db        448 ESKSFLEDMDKKSLETKSLEMTFIPTTED----TGKVLVCLAKLHIDEMEFEPKQRQS-- 501

Qy        297 LTLALEAKTGKLHQEVNLVVMRATQLQK----NLTCEVWGPTSPKLMLSLKLENKEAKVS 352
              | |          ::  :||  :: :::    |:||   |  :| :: | :| |   : |
Db        502 -TQTLYVNVAP--RDTTVVVSPSSIVEEGSPVNMTCSSDGLPAPNILWSRRLSNGRLQ-S 557

Qy        353 KREKPVWVLN----PEAGMWQC 370
              | |:  |       ::|:: |
Db        558 LSEDPILTLTSAKMEDSGIYVC 579


RESULT 30
Q9GKR2
ID   Q9GKR2      PRELIMINARY;      PRT;   650 AA.
AC   Q9GKR2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vascular cell adhesion molecule-1 6D variant lacking D7
DE   precursor.
GN   VCAM1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Holstein; TISSUE=Muscle;
RA   Muroya S., Nakajima I., Chikuni K.;
RT   "Novel alternative splicing variant of bovine vascular cell adhesion
RT   molecule-1.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; AB052747; BAB19782.1; -.
DR   HSSP; P19320; 1VCA.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 4.
KW   Immunoglobulin domain; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    650       VASCULAR CELL ADHESION MOLECULE-1 6D
FT                              VARIANT LACKING D7.
SQ   SEQUENCE   650 AA;  71821 MW;  24244573EF2B5E40 CRC64;

  Query Match             7.1%;  Score 165;  DB 6;  Length 650;
  Best Local Similarity   21.7%;  Pred. No. 0.00044;
  Matches   88;  Conservative   69;  Mismatches  131;  Indels  118;  Gaps   19;

Qy         32 KKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGN 91
              ::||:| :|| ::   : :  |                     ||           | ||
Db        238 QEGDSVTMTCASAGLPAPRILWS------------------KKL----------DNGN 267

Qy         92 FPLIIKN-------LKIEDSDTYICE----VEDQKEEVQLLV----FGLTANSDTHLLQ- 135
                |: :|       :: |||  |:||     |   ::||:| |     | :  :    :
Db        268 RQLLSENATLTLISMRAEDSGIYVCEGNNPVGKDRKEVKLTVQEKNFTVEISPGPQIAAQ 327

Qy        136 -GQSLTLTLESPPGSSPSVQCR----SPRGKNI--QGGK-TLSVSQLELQDSGTWTCTVL 187
               | |: || :     |||   |    ||   |:  :| | ||::|  :  ::    : |||:
Db        328 VGDSVVLTCDVRDCESPSFSWRTLIDSPLNGNVRSEGSKSTLTLSPVSFENEHFYLCTVM 387

Qy        188 QNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPL--------AFTVEKLTGSGELW 239
               |||:| :| :   :|             ::| | ||        :  |  :  | :|
Db        388 CGQKKLEKRIQVKPYSF---------PSNPEIEMSGPLVSGNPVTVSCRVPNVYPSDQLE 438

Qy        240 WQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTL 299
               :  :  |      ||   :|:|| |:  :   |:              :|     ||: :
Db        439 IELFKGESIMMNKTF---SKDVS-KKALETKSLE----------KTFIPTTEDIGNVLV 483

Qy        300 AL--------------EAKTGKLH-----QEVNLVVMRATQLQK----NLTCEVWGPTSP 336
               |              |  |:    ::  ::|  :: |::    |:||   |  :|
Db        484 CLARLPIDEMEFELKQRQSTQTLYVNVAPRDTTILVSPSSILEEGSSVNMTCSSNGLPAP 543

Qy        337 KLMLSLKLENKEAK-VSKREKPVWVLNPEAGMWQCLLSDSGQVLLE 381
               |:: | :| | : : :|:        |         : |||  : |
Db        544 KILWSRQLSNGDLQPISE--------NATLSFMSTKMEDSGIYVCE 581


RESULT 31
Q9GKR3
ID   Q9GKR3      PRELIMINARY;      PRT;   739 AA.
AC   Q9GKR3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vascular cell adhesion molecule-1 7D variant precursor.
GN   VCAM1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Holstein; TISSUE=Muscle;
RA   Muroya S., Nakajima I., Chikuni K.;
RT   "Novel alternative splicing variant of bovine vascular cell adhesion
RT   molecule-1.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB052746; BAB19650.1; -.
DR   HSSP; P19320; 1VCA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016337; P:cell-cell adhesion; IEA.
DR   InterPro; IPR003987; ICAM_VCAM-1.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003989; VCAM-1.
DR   Pfam; PF00047; ig; 7.
DR   PRINTS; PR01472; ICAMVCAM1.
DR   PRINTS; PR01474; VCAM1.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 5.
KW   Immunoglobulin domain; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    739       VASCULAR CELL ADHESION MOLECULE-1 7D
FT                              VARIANT.
SQ   SEQUENCE   739 AA;  81398 MW;  CA32E3141117A1BF CRC64;

  Query Match             7.1%;  Score 165;  DB 6;  Length 739;
  Best Local Similarity   21.7%;  Pred. No. 0.00053;
  Matches   88;  Conservative   69;  Mismatches  131;  Indels  118;  Gaps   19;

Qy         32 KKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGN 91
              ::||:| :|| ::   : :  |                     ||           | ||
Db        238 QEGDSVTMTCASAGLPAPRILWS------------------KKL----------DNGN 267

Qy         92 FPLIIKN-------LKIEDSDTYICE----VEDQKEEVQLLV----FGLTANSDTHLLQ- 135
                |: :|       :: |||  |:||     |   ::||:| |     | :  :    :
Db        268 RQLLSENATLTLISMRAEDSGIYVCEGNNPVGKDRKEVKLTVQEKNFTVEISPGPQIAAQ 327

Qy        136 -GQSLTLTLESPPGSSPSVQCR----SPRGKNI--QGGK-TLSVSQLELQDSGTWTCTVL 187
               | |: || :     |||   |    ||   |:  :| | ||::|  :  ::    : |||:
Db        328 VGDSVVLTCDVRDCESPSFSWRTLIDSPLNGNVRSEGSKSTLTLSPVSFENEHFYLCTVM 387

Qy        188 QNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPL--------AFTVEKLTGSGELW 239
               |||:| :| :   :|             ::| | ||        :  |  :  | :|
Db        388 CGQKKLEKRIQVKPYSF---------PSNPEIEMSGPLVSGNPVTVSCRVPNVYPSDQLE 438

Qy        240 WQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTL 299
               :  :  |      ||   :|:|| |:  :   |:              :|     ||: :
Db        439 IELFKGESIMMNKTF---SKDVS-KKALETKSLE----------KTFIPTTEDIGNVLV 483

Qy        300 AL--------------EAKTGKLH-----QEVNLVVMRATQLQK----NLTCEVWGPTSP 336
               |              |  |:    ::  ::|  :: |::    |:||   |  :|
Db        484 CLARLPIDEMEFELKQRQSTQTLYVNVAPRDTTILVSPSSILEEGSSVNMTCSSNGLPAP 543

Qy        337 KLMLSLKLENKEAK-VSKREKPVWVLNPEAGMWQCLLSDSGQVLLE 381
               |:: | :| | : : :|:        |         : |||  : |
Db        544 KILWSRQLSNGDLQPISE--------NATLSFMSTKMEDSGIYVCE 581


RESULT 32
Q63669
ID   Q63669      PRELIMINARY;      PRT;   739 AA.
AC   Q63669;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vascular cell adhesion molecule 1 precursor.
GN   VCAM-1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague Dawley; TISSUE=Lung;
RX   MEDLINE=92305064; PubMed=1377031;
RA   Williams A.J., Atkins R.C., Fries J.W., Gimbrone M.A.Jr.,
RA   Cybulsky M.I., Collins T.;
RT   "Nucleotide sequence of rat vascular cell adhesion molecule-1 cDNA.";
RL   Biochim. Biophys. Acta 1131:214-216(1992).
DR   EMBL; X63722; CAA45254.1; -.
DR   HSSP; P19320; 1VCA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016337; P:cell-cell adhesion; IEA.
DR   InterPro; IPR003987; ICAM_VCAM-1.
DR   InterPro; IPR007110; Ig-like.
```

```
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003989; VCAM-1.
DR   Pfam; PF00047; ig; 5.
DR   PRINTS; PR01472; ICAMVCAM1.
DR   PRINTS; PR01474; VCAM1.
DR   SMART; SM00408; IGc2; 4.
DR   PROSITE; PS50835; IG_LIKE; 5.
KW   Immunoglobulin domain; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    739       VASCULAR CELL ADHESION MOLECULE 1.
SQ   SEQUENCE   739 AA;  81136 MW;  A5AAD1172F67FB96 CRC64;

  Query Match             7.0%;  Score 162;  DB 11;  Length 739;
  Best Local Similarity   22.0%;  Pred. No. 0.00089;
  Matches   86;  Conservative   60;  Mismatches  155;  Indels   90;  Gaps   14;

Qy         33 KGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNF 92
              :|  | :|| :    : :  |                |  |  :|            ||
Db        238 EGAAVTMTCASEGLPAPEIFWSKK-----------LDNGVLQLL----------SGNA 274

Qy         93 PLIIKNLKIEDSDTYICE----VEDQKEEVQLLV------FGLTANSDTHLLQGQSLTLT 142
               |  :  :::|||  |:||    |   | ||:|:|       ::  |          | |: ||
Db        275 TLTLIAMRMEDSGIYVCEGVNLVGRDKTEVELIVQEKPFTVDISPGSQVAAQVGDSVVLT 334

Qy        143 LESPPGSSPSVQCRSPRGKNIQG-------GKTLSVSQLELQDSGTWTCTVLQNQKKVEF 195
                :    |||   |:     : |         ||:|: : ::|  :: |||   ::|:|
Db        335 CAAVGCDSPSFSWRTQTDSPLNGEVRDEGATSTLTLSPVGVEDEHSYLCTVTCQRRKLEK 394

Qy        196 KIDIVVLAFQKASSIVYKKEGEQVEFSFPLA-------------------FTVEKLTGSG 236
              | : | :|           |  ::| | ||                       :| | |
Db        395 TIQVEVYSF---------PEDPEIEISGPLVHGRPVTVNCTVPNVYPFDHLEIELLKGET 445

Qy        237 ELWWQAERAS-SSKSWITFDLKNKEVSVKRVTQDPKLQMGKKL----PLHLTLPQALP-Q 290
              | : |    :||    : |: |::     :|     || |     || :  :: | |
Db        446 TLLNKFLREEIGTKS---LETKSLEMTFIPTAED----TGKALVCLAKLHSSQMESEPKQ 498

Qy        291 YAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAK 350
                :  | : :  |   :    : |    : :  ||||   |  :||:: | :|:| | :
Db        499 RQSTQTLYVNVAPKEPTIWVSPSPVPEEGSPV--NLTCSSDGFPTPKILWSRQLKNGELQ 556

Qy        351 VSKREKPVWVLNPEAGMWQCLLSDSGQVLLE 381
                :       |          : |||  : |
Db        557 PLSQ-------NTTLSFMATKMEDSGIYVCE 580


RESULT 33
Q29123
ID   Q29123      PRELIMINARY;      PRT;   538 AA.
AC   Q29123;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vascular cell adhesion molecule.
GN   VCAM.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=94271236; PubMed=7516159;
RA   Tsang Y.T., Haskard D.O., Robinson M.K.;
RT   "Cloning and expression kinetics of porcine vascular cell adhesion
RT   molecule.";
RL   Biochem. Biophys. Res. Commun. 201:805-812(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=96106491; PubMed=8525525;
RA   Mueller J.P., Evans M.J., Cofiell R., Rother R.P., Matis L.A.,
RA   Elliott E.A.;
RT   "Porcine vascular cell adhesion molecule (VCAM) mediates endothelial
RT   cell adhesion to human T cells. Development of blocking antibodies
RT   specific for porcine VCAM.";
RL   Transplantation 60:1299-1306(1995).
DR   EMBL; L43124; AAB59281.1; -.
DR   HSSP; P19320; 1VCA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016337; P:cell-cell adhesion; IEA.
DR   InterPro; IPR003987; ICAM_VCAM-1.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003989; VCAM-1.
DR   Pfam; PF00047; ig; 5.
DR   PRINTS; PR01472; ICAMVCAM1.
DR   PRINTS; PR01474; VCAM1.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 4.
KW   Immunoglobulin domain.
SQ   SEQUENCE   538 AA;  58713 MW;  8A7CD36D0A2F0717 CRC64;

  Query Match             6.9%;  Score 161.5;  DB 6;  Length 538;
  Best Local Similarity   23.7%;  Pred. No. 0.00063;
  Matches   66;  Conservative   49;  Mismatches   90;  Indels   73;  Gaps   11;

Qy         32 KKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGN 91
              ::||:| :|||:    : |  |                ||  |  | :      ||
Db        235 QEGDSVMMTCTSEGLPAPQISW-------------------SKKLDNGDQQLL---SGN 271

Qy         92 FPLIIKNLKIEDSDTYICE----VEDQKEEVQLLV------FGLTANSDTHLLQGQSLTL 141
               |  :  :::|||  |:||    |   ::||:| |         ::  |  : | :| |: :
Db        272 ATLTLIAMRMEDSGIYVCEGVNPVGTNRKEVELTVQVAPRDTTISVNPSSTLEEGSSVNM 331

Qy        142 TLESPPGSSPSV----QCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ----NQKKV 193
                |  |    :| :    : |    : :    ||:::  :::||| : |  :     |:|:|
Db        332 TCSSDGFPAPKILWSKKLRDGNLEPLSENTTLTLTSTKMEDSGIYVCDGINQAGINRKEV 391

Qy        194 EFKI-----DIVVLAFQKAS-------------------SIVYKKE---GEQVEFSFPLA 226
              |  |     |: ::||  |                    |: ||:   |: |  |   |
Db        392 ELIIQAAPKDLQLIAFPSESVKEGDTVIISCTCGNVPPTLIILKKKAETGDTVLKSTDGA 451

Qy        227 FTVEKL----TGSGELWWQAERASSSKSWITFDLKNKE 260
              :|: :      |  |   : |     :| || |:| :|
Db        452 YTIHRARLADAGVYECESKNEIGLQLRS-ITLDVKGRE 488


RESULT 34
Q8NFP4
ID   Q8NFP4      PRELIMINARY;      PRT;   955 AA.
AC   Q8NFP4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glycosyl-phosphatidyl-inositol-MAM.
GN   GPIM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22077705; PubMed=12082541;
RA   De Juan C., Iniesta P., Gonzalez-Quevedo R., Moran A.,
RA   Sanchez-Pernaute A., Torres A.J., Balibrea J.L., Diaz-Rubio E.,
RA   Cruces J., Benito M.;
RT   "Genomic organization of a novel glycosylphosphatidylinositol MAM gene
RT   expressed in human tissues and tumors.";
RL   Oncogene 21:3089-3094(2002).
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR   EMBL; AF478693; AAM77220.1; -.
DR   Genew; HGNC:19267; MDGA1.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016163; F:nitrogenase activity; IEA.
DR   GO; GO:0009399; P:nitrogen fixation; IEA.
DR   InterPro; IPR008957; FN_III-like.
```

```
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000998; MAM_domain.
DR   InterPro; IPR000318; Nitrognse_comp1.
DR   Pfam; PF00047; ig; 6.
DR   Pfam; PF00629; MAM; 1.
DR   PRINTS; PR00020; MAMDOMAIN.
DR   SMART; SM00409; IG; 6.
DR   SMART; SM00408; IGc2; 6.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS50835; IG_LIKE; 6.
DR   PROSITE; PS50060; MAM_2; 1.
DR   PROSITE; PS00699; NITROGENASE_1_1; 1.
KW   Glycoprotein; Immunoglobulin domain.
SQ   SEQUENCE   955 AA;  105790 MW;  BD41A1EB10A05962 CRC64;

  Query Match             6.8%;  Score 157;  DB 4;  Length 955;
  Best Local Similarity   23.3%;  Pred. No. 0.0031;
  Matches  112;  Conservative   64;  Mismatches  181;  Indels  124;  Gaps   25;

Qy         36 TVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI 95
Db        152 TVFLRCTVNSNPPARFIWKRGSD-----------TLSHSQDNG-VDIYEPLYTQGETKVL 199

Qy         96 -IKNLKIEDSDTYICEVEDQKE---EVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSP 151
Db        200 KLKNLRPQDYASYTCQVSVRNVCGIPDKAITFRLTNTTAPPALK-LSVNETLVVNPGENV 258

Qy        152 SVQC-----------------RSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN---- 189
Db        259 TVQCLLTGGDPLPQLQWSHGPGPLPLGALAQGG-TLSIPSVQARDSGYYNCTATNNVGNP 317

Qy        190 QKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFSFPL-------AFTVEKLTGSGELW 239
Db        318 AKKT---VNLLVRSMKNATFQITPDVIKESENIQLGQDLKLSCHVDAVPQEKVTYQ---W 371

Qy        240 WQ-AERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHL----------TLPQA- 287
Db        372 FKNGKPARMSKRLL---VTRNDPELPAVTSSLEL-----IDLHFSDYGTYLCMASFPGAP 423

Qy        288 LPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENK 347
Db        424 VPDLSVEVNISSETVPPTISVPKGRAVVTVREGS-PAELQCEVRGKPRPPVLWS--RVDK 480

Qy        348 EAK-----VSKREKPVWVLNPE------AGMWQC---------LLSDSGQVLLE------ 381
Db        481 EAALLPSGLPLEETPDGKLRLERVSRDMSGTYRCQTARYNGFNVRPREAQVQLNVQFPPE 540

Qy        382 ---SNIKVLPTWSTPVHPRASALPAPP----------TGSALPDPQTASALPDPPAASAL 428
Db        541 VEPSSQDVRQALGRPVLLRCSLLRGSPQRIASAVWRFKGQLLPPP------PVVPAAAEA 594

Qy        429 P 429
Db        595 P 595


RESULT 35
Q28939
ID   Q28939      PRELIMINARY;      PRT;   538 AA.
AC   Q28939;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vascular cell adhesion molecule precursor.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Aorta;
RX   MEDLINE=94271236; PubMed=7516159;
RA   Tsang Y.T., Haskard D.O., Robinson M.K.;
RT   "Cloning and expression kinetics of porcine vascular cell adhesion
RT   molecule.";
RL   Biochem. Biophys. Res. Commun. 201:805-812(1994).
DR   EMBL; U08351; AAA21542.1; -.
DR   PIR; JC2457; JC2457.
DR   HSSP; P19320; 1VCA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016337; P:cell-cell adhesion; IEA.
DR   InterPro; IPR003987; ICAM_VCAM-1.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003989; VCAM-1.
DR   Pfam; PF00047; ig; 5.
DR   PRINTS; PR01472; ICAMVCAM1.
DR   PRINTS; PR01474; VCAM1.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 4.
KW   Immunoglobulin domain; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    538       VASCULAR CELL ADHESION MOLECULE.
SQ   SEQUENCE   538 AA;  58795 MW;  EC29D11B224F7261 CRC64;

  Query Match             6.7%;  Score 155.5;  DB 6;  Length 538;
  Best Local Similarity   23.4%;  Pred. No. 0.0018;
  Matches   65;  Conservative   49;  Mismatches   91;  Indels   73;  Gaps   11;

Qy         32 KKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGN 91
Db        235 QEGDSMMMTCTSEGLPAPQISW-------------------SKKLDNGDQQLL---SGN 271

Qy         92 FPLIIKNLKIEDSDTYICE----VEDQKEEVQLLV------FGLTANSDTHLLQGQSLTL 141
Db        272 ATLTLIAMRMEDSGIYVCEGVNPVGTNRKEVELTVQVAPRDTTISVNPSSTLEEGSSVNM 331

Qy        142 TLESPPGSSPSV----QCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ----NQKKV 193
Db        332 TCSSDGFPAPKILWSKKLRDGNLEPLSENTTLTLTSTKMEDSGIYVCEGINQAGINRKEV 391

Qy        194 EFKI-----DIVVLAFQKAS-------------------SIVYKKE---GEQVEFSFPLA 226
Db        392 ELIIQAAPKDLQLTAFPSESVKEGDTVIISCTCGNVPPTLIILKKKAETGDTVLKSTDGA 451

Qy        227 FTVEKL----TGSGELWWQAERASSSKSWITFDLKNKE 260
Db        452 YTIHRARLADAGVYECESKNEIGLQLRS-ITLDVKGRE 488


RESULT 36
Q9UDE5
ID   Q9UDE5      PRELIMINARY;      PRT;    31 AA.
AC   Q9UDE5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CD4=55 kDa glycoprotein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92072595; PubMed=1961196;
RA   Lederman S., DeMartino J.A., Daugherty B.L., Foeldvari I.,
RA   Yellin M.J., Cleary A.M., Berkowitz N., Lowy I., Braunstein N.S.,
RA   Mark G.E.;
RT   "A single amino acid substitution in a common African allele of the
RT   CD4 molecule ablates binding of the monoclonal antibody, OKT4.";
RL   Mol. Immunol. 28:1171-1181(1991).
DR   HSSP; P01730; 1WIO.
DR   InterPro; IPR007110; Ig-like.
FT   NON_TER       1      1
```

```
FT   NON_TER      31     31
SQ   SEQUENCE   31 AA;  3658 MW;  A7E9C61F5DCF6CEF CRC64;

  Query Match             6.5%;  Score 152;  DB 4;  Length 31;
  Best Local Similarity   96.8%;  Pred. No. 6.3e-05;
  Matches   30;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        250 SWITFDLKNKEVSVKRVTQDPKLQMGKKLPL 280
              ||||||||||||||| ||||||||||||||
Db          1 SWITFDLKNKEVSVKWVTQDPKLQMGKKLPL 31


RESULT 37
Q92626
ID   Q92626      PRELIMINARY;      PRT;  1496 AA.
AC   Q92626;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MYELOBLAST KIAA0230 (Fragment).
GN   KIAA0230.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.,
RA   Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=95048383; PubMed=7959781;
RA   Weiler S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RA   Trent J.M.;
RT   "Assignment of a human melanoma associated gene MG50 (D2S448) to
RT   chromosome 2p25.3 by fluorescence in situ hybridization.";
RL   Genomics 22:243-244(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RA   Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT   "Identification of a novel melanoma gene (MG50) - likely the gene for
RT   IL-1 receptor antagonist - which encodes epitopes recognized by human
RT   cytolytic T lymphocytes.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D86983; BAA13219.1; -.
DR   EMBL; AF200348; AAF06354.1; -.
DR   HSSP; P05164; 1CXP.
DR   GO; GO:0005152; F:interleukin-1 receptor antagonist activity; NAS.
DR   GO; GO:0006955; P:immune response; NAS.
DR   InterPro; IPR002007; Anim_peroxidase.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR002016; Peroxidase.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF03098; An_peroxidase; 1.
DR   Pfam; PF00047; ig; 4.
DR   Pfam; PF00560; LRR; 5.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF00093; vwc; 1.
DR   PRINTS; PR00457; ANPEROXIDASE.
DR   SMART; SM00408; IGc2; 4.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 4.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS50835; IG_LIKE; 4.
DR   PROSITE; PS50292; PEROXIDASE_3; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Immunoglobulin domain.
FT   NON_TER      1      1
SQ   SEQUENCE   1496 AA;  167209 MW;  E9B9A7069BF1ABFF CRC64;

  Query Match             6.5%;  Score 152;  DB 4;  Length 1496;
  Best Local Similarity   21.5%;  Pred. No. 0.014;
  Matches  100;  Conservative   57;  Mismatches  155;  Indels  154;  Gaps   19;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |:||  || |      :  |  :|                ::|: : ||| :| | |
Db        277 GNTVYFTCRAEGNPKPEIIWLRNN---------------NELSMKTDSRLNLLDDGT-- 318

Qy         94 LIIKNLKIEDSDTYICEV-----EDQKEEVQLLVFGLTAN-------SDTHLLQGQSLTL 141
              |:|:| :  |   | |       | : :|| |  ||  |           :| :| |:|:||
Db        319 LMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTL 378

Qy        142 ----TLESPPGSSPSVQCRSP----RGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
                  |   ||  | :   |:|       ||      | :  :   ||| : |:   |  |
Db        379 ECSATGHPPPRISWTRGDRTPLPVDPRVNITPSGGLYIQNVVQGDSGEYACSATNNIDSV 438

Qy        194 EFKIDIVVLAFQKASSIVYKK---EGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKS 250
                 |:| |  : :     :   ||: |:|                    | |   :
Db        439 HATAFIIVQALPQFTVTPQDRVVIEGQTVDF------------------QCEAKGNPPP 479

Qy        251 WITFDLKNKEVSVKR---VTQDPKLQM--------------------GKKLPLHLTL-PQ 286
                | :     ::|| |   |     |::                        :|:  |||: |:
Db        480 VIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPR 539

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLEN 346
                | :|   : |             ||   |         | |   |   | :       |
Db        540 VTPVFASIPSDTTV----------EVGANV--------QLPCSSQGEPEPAITW-----N 576

Qy        347 KEAKVSKREKPVWVLNPE------------AGMWQCLL-----SDSGQVLLESNIKVLPT 389
              |:  |   |   : ::||             || ::|:       | |  ::|  |:
Db        577 KDG-VQVTESGKFHISPEGFLTINDVGPADAGRYECVARNTIGSASVSMVLSVNV----- 630

Qy        390 WSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVI 435
                                ||   |   ||  |::: |:| :
Db        631 ---------------------PD---VSRNGDPFVATSIVEAIATV 652


RESULT 38
Q865F2
ID   Q865F2      PRELIMINARY;      PRT;   739 AA.
AC   Q865F2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Adhesion molecule VCAM-1.
GN   VCAM-1.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schnurr K., Banning A., Kupper D., Muller-Schmehl K.,
RA   Brigelius-Flohe R.;
RT   "Modulation of basal and interleukin-1-induced adhesion molecule
RT   expression by phospholipid hydroperoxide glutathione peroxidase and
RT   15-lipoxygenase in rabbit aortic smooth muscle cells.";
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY212510; AAO52742.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016337; P:cell-cell adhesion; IEA.
```

```
DR   InterPro; IPR003987; ICAM_VCAM-1.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003989; VCAM-1.
DR   Pfam; PF00047; ig; 5.
DR   PRINTS; PR01472; ICAMVCAM1.
DR   PRINTS; PR01474; VCAM1.
DR   SMART; SM00409; IG; 5.
DR   SMART; SM00408; IGc2; 5.
DR   PROSITE; PS50835; IG_LIKE; 5.
SQ   SEQUENCE   739 AA;  81806 MW;  7AD0D32511E19342 CRC64;

  Query Match             6.5%;  Score 151.5;  DB 6;  Length 739;
  Best Local Similarity   19.9%;  Pred. No. 0.0057;
  Matches   83;  Conservative   71;  Mismatches  149;  Indels  115;  Gaps   17;

Qy         32 KKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGN 91
Db        237 QEGGSVTMTCSSEGLPVPEIFW--------------------SKKQDNGNLQRL---SGN 273

Qy         92 FPLIIKNLKIEDSDTYICEVEDQ----KEEVQLLVFGLTANSDTHLLQGQSLTLTLESPP 147
Db        274 ATLTLIAMRMEDSGIYVCEGVNQIGKSRKEVELIV--------------QEKPFTVEISP 319

Qy        148 G--------------------SSPSVQCRSPRGKNIQGGKT-------LSVSQLELQDSG 180
Db        320 GPRIAAQIGDPVVLTCSVRGCETPSFSWRTQIDSPLNGQVTSEGTKSLLTLSPVSFENEH 379

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFS------FPL-AFTVE 230
Db        380 SYLCTVTCGHKKLEKGIQVELYSFPRDPEIELSGPPVNGRPVTVSCKVPNVYPFDXLEIE 439

Qy        231 KLTGSGELWWQAERASSSKSWI-TFDLKNKEVSVKRVTQDPKLQ-MGKKL----PLHL-- 282
Db        440 LLKG--------ETMMKNKEFLEEEDKKSLETKSLEMTFIPTMEDTGKVLVCQAKLHIDE 491

Qy        283 ------TLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSP 336
Db        492 MEFEPKQRQSTQPLFVNVAPRDIAVWVSPSSIVEEGRSV---------NMTCSSYGLPAP 542

Qy        337 KLMLSLKLENKEAKVSKREKPVWVLN---PEAGMWQC---LLSDSGQVLLESNIKVLP 388
Db        543 KILWSRQLKNGDLQPLSENTTLALISTKLEDSGIYVCEGINLAGKSRKEVELVIQVAP 600


RESULT 39
Q80U60
ID   Q80U60      PRELIMINARY;      PRT;  1431 AA.
AC   Q80U60;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MKIAA0230 protein (Fragment).
GN   MKIAA0230.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22579291; PubMed=12693553;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA   Nakajima D., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT   II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT   cDNAs identified by screening of terminal sequences of cDNA clones
RT   randomly sampled from size-fractionated libraries.";
RL   DNA Res. 10:35-48(2003).
DR   EMBL; AK122223; BAC65505.1; -.
DR   GO; GO:0004601; F:peroxidase activity; IEA.
DR   GO; GO:0006979; P:response to oxidative stress; IEA.
DR   InterPro; IPR002007; Anim_peroxidase.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR007092; LRR_SDS22.
DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR002016; Peroxidase.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF03098; An_peroxidase; 1.
DR   Pfam; PF00047; ig; 4.
DR   Pfam; PF00560; LRR; 5.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF00093; vwc; 1.
DR   PRINTS; PR00457; ANPEROXIDASE.
DR   SMART; SM00409; IG; 4.
DR   SMART; SM00408; IGc2; 4.
DR   SMART; SM00406; IGv; 3.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00366; LRR_PS; 4.
DR   SMART; SM00369; LRR_TYP; 5.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS50835; IG_LIKE; 4.
DR   PROSITE; PS50292; PEROXIDASE_3; 1.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   1431 AA;  160591 MW;  6BA952436DA54B72 CRC64;

  Query Match             6.2%;  Score 145;  DB 11;  Length 1431;
  Best Local Similarity   20.4%;  Pred. No. 0.045;
  Matches   96;  Conservative   59;  Mismatches  153;  Indels  162;  Gaps   20;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
Db        213 GNTVYFTCRAEGNPKPEIIWLRNN----------------NELSMKTDSRLNLLDDGT-- 254

Qy         94 LIIKNLKIEDSDTYICEV-----EDQKEEVQLLVFGLTAN-------SDTHLLQGQSLTL 141
Db        255 LMIQNTQEADEGVYQCMAKNVAGEAKTQEVTLRYLGSPARPTFVIQPQNTEVLVGES--V 312

Qy        142 TLESPPGSSPSVQCRSPRGK----------NIQGGKTLSVSQLELQDSGTWTCTVLQNQK 191
Db        313 TLECSATGHPLPQITWTRGDRTPLPIDPRVNITPSGGLYIQNVAQSDSGEYTCFASNSVD 372

Qy        192 KVEFKIDIVVLAFQK-----ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERAS 246
Db        373 SIHATAFIIVQALPQFTVTPQSRVVI--EGQTVDF------------------QCAAKG 411

Qy        247 SSKSWITFDLKNKEVSVKR---VTQDPKLQM-------------------GKKLPLHLT 283
Db        412 HPQPVIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLT 471

Qy        284 L-PQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSL 342
Db        472 VQPRVTPVFASIPS---------------DMTVEVGTNVQ--LPCSSQGEPEPAITW-- 511

Qy        343 KLENKEAKVSKREKPVWVLNPE------------AGMWQCLLSD-----SGQVLLESNIK 385
Db        512 ---NKDG-VQVTESGKFHISPEGFLTINDVGTADAGRYECVARNTIGYASVSMVLSVNV- 566

Qy        386 VLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVI 435
Db        567 ------------------------PD---VSRNGDPYVATSIVEAIATV 588


RESULT 40
Q8IZP8
ID   Q8IZP8      PRELIMINARY;      PRT;   437 AA.
AC   Q8IZP8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nectin-like protein 3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gingrich J.R., D'Angelo A., Chang G.M., Greenberg N.M.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF538973; AAN16368.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 2.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 3.
SQ   SEQUENCE   437 AA;  47593 MW;  D23DB347F5E5F023 CRC64;

  Query Match             6.2%;  Score 144.5;  DB 4;  Length 437;
  Best Local Similarity   19.0%;  Pred. No. 0.0095;
  Matches   91;  Conservative   73;  Mismatches  166;  Indels  149;  Gaps   18;

Qy         10 LLLVLQLALLPAATQGNKV------------VLGKKGDTVELTCTASQKKSIQFHWKNSN 57
              :| :  |:|:|||   |||             |   :| |  |||   |  :    | |
Db          3 VLFLCNLSLVPAAASKNKVKGSQGQFPLTQNVTVVEGGTAILTCRVDQNDNTSLQWSNPA 62

Qy         58 QIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICE---VEDQ 114
              |     | :    :  ::| :  |: | :    : : :: : |   | |    :  :
Db         63 Q------QTLYFDDKKALRDNRIELVRASWHE--LSISVSDVSLSDEGQYTCSLFTMPVK 114

Qy        115 KEEVQLLVFGL-----TANSDTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKT 168
              :  | | |:      :   : :::|  : || ::  || |:   |     | |:
Db        115 TSKAYLTVLGVPEKPQISGFSSPVMEGDLMQLTCKT-SGSKPAADIRWFKNDKEIK---- 169

Qy        169 LSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFT 228
              |  |: :|:   | ||      ::|::|          ::: : : | :  :  :|
Db        170 -DVKYLKKEDANRKTFTV---SSTLDFRVD----RSDDGVAVICRVDHESLNTTPQVAMQ 221

Qy        229 VEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQAL 288
              | ::                :   ||| :  |  | |: |
Db        222 VLEI-------------------------HYTPSVKIIPSTPFPQEGQPL---------- 246

Qy        289 PQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLK----- 343
                                                ||||  |   |: :|  |
Db        247 ------------------------------------ILTCESKGKPLPEPVLWTKDGGEL 270

Qy        344 LENKEAKVSKREKPVWVLN-PEAGMWQC--------------LLSDSGQVLLESNIKVL 387
              :     || ||  :  ||  : | ::|               :: |    || : |  :
Db        271 PDPDRMVVSGRELNILFLNKTDNGTYRCEATNTIGQSSAEYVLIVHDVPNTLLPTTI--I 328

Qy        388 PTWSTPVHPRASALPAPPTGSALPDPQTASALPDP--------PAASALPAALAVISFL 438
              |: :|       |:   || ||     | |:: ||        |  : :   :||: |:
Db        329 PSLTTATVTTTVAITTSPTTSA-----TTSSIRDPNALAGQNGPDHALIGGIVAVVVFV 382


RESULT 41
Q91X98
ID   Q91X98      PRELIMINARY;      PRT;   760 AA.
AC   Q91X98;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   VCAM1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011159; AAH11159.1; -.
DR   MGD; MGI:98926; Vcam1.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016337; P:cell-cell adhesion; IEA.
DR   InterPro; IPR003987; ICAM_VCAM-1.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003989; VCAM-1.
DR   Pfam; PF00047; ig; 5.
DR   PRINTS; PR01472; ICAMVCAM1.
DR   PRINTS; PR01474; VCAM1.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 5.
KW   Hypothetical protein; Immunoglobulin domain.
FT   NON_TER       1       1
SQ   SEQUENCE   760 AA;  83862 MW;  9EF64744BB484553 CRC64;

  Query Match             6.2%;  Score 144;  DB 11;  Length 760;
  Best Local Similarity   20.2%;  Pred. No. 0.022;
  Matches   79;  Conservative   65;  Mismatches  147;  Indels  100;  Gaps   15;

Qy         32 KKGDTVELTCTASQKKSIQFHW--KNSNQI-KILGNQGSFLTKGPSKLNDRADSRRSLWD 88
              ::|  | :||::    : :  |  |  |:: ::|
Db        258 QEGGAVTMTCSSEGLPAPEIFWGRKLDNEVLQLL------------------------- 291

Qy         89 QGNFPLIIKNLKIEDSDTYICE----VEDQKEEVQLLV------FGLTANSDTHLLQGQS 138
              ||  | :  :::|||  |:||    :   | ||:|:|       ::  |      | |
Db        292 SGNATLTLIAMRMEDSGVYVCEGVNLIGRDKAEVELVVQEKPFIVDISPGSQVAAQVGDS 351

Qy        139 LTLTLESPPGSSPSVQCRSPRGKNIQG-------GKTLSVSQLELQDSGTWTCTVLQNQK 191
              : ||  :   |||   |:     : |         || :|  :  :|  ::  | |   |:
Db        352 VVLTCAAIGCDSPSFSWRTQTDSPLNGVVRNEGAKSTLVLSSVGFEDEHSYLCAVTCLQR 411

Qy        192 KVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLA------------------FTVEKL 232
              :| :  : | :|          |   :: | ||                      :| |
Db        412 TLEKRTQVEVYSF---------PEDPVIKMSGPLVHGRPVTVNCTVPNVYPFDHLEIELL 462

Qy        233 TGSGELWWQAERASSSKSWI-TFDLKNKEVSVKRVTQDPKLQ-MGKKL----PLHLTLPQ 286
              |       |     | ::    :|: |  :   |  | ::  || |     ||    :
Db        463 KG--------ETTLMKKYFLEEMGIKSLETKILETTFIPTIEDTGKSLVCLARLHSGEME 514

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQK----NLTCEVWGPTSPKLMLSL 342
              : |:   |          |     :|  : |  :  |::    ||||   |  :||:: |
Db        515 SEPKQRQSVQPLYVNVAP-----KETTIWVSPSPILEEGSPVNLTCSSDGIPAPKILWSR 569

Qy        343 KLENKEAKVSKREKPVWVLN---PEAGMWQC 370
              :| | | :       :  ::    ::|:: |
Db        570 QLNNGELQPLSENTTLTFMSTKRDDSGIYVC 600


RESULT 42
Q96RW7
ID   Q96RW7      PRELIMINARY;      PRT;  5636 AA.
AC   Q96RW7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemicentin.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Trent J.;
RT   "Human hemicentin gene.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF156100; AAK68690.1; -.
DR   GO; GO:0005727; C:extrachromosomal circular DNA; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
```

```
DR   GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000875; Cecropin.
DR   InterPro; IPR001434; DUF11.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000169; SHprot_acsite.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF00047; ig; 44.
DR   Pfam; PF00090; tsp_1; 6.
DR   SMART; SM00179; EGF_CA; 7.
DR   SMART; SM00408; IGc2; 43.
DR   SMART; SM00209; TSP1; 6.
DR   SMART; SM00327; VWA; 1.
DR   TIGRFAMs; TIGR01451; B_ant_repeat; 9.
DR   PROSITE; PS00010; ASX_HYDROXYL; 5.
DR   PROSITE; PS00268; CECROPIN; 1.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01187; EGF_CA; 8.
DR   PROSITE; PS50835; IG_LIKE; 44.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR   PROSITE; PS50092; TSP1; 6.
KW   EGF-like domain; Immunoglobulin domain.
SQ   SEQUENCE   5636 AA;  613660 MW;  F000B319CED7B52C CRC64;

  Query Match             6.2%;  Score 143;  DB 4;  Length 5636;
  Best Local Similarity   20.4%;  Pred. No. 0.44;
  Matches   93;  Conservative   56;  Mismatches  168;  Indels  138;  Gaps   21;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
Db        807 GSNVTLPCYVQGYPEPTIKWRRLDNMPIFSR--PFSVSSISQLRTGA------------- 851

Qy         94 LIIKNLKIEDSDTYICEVEDQ----KEEVQLLVFGLTA------NSDTHLLQGQSLTLTL 143
Db        852 LFILNLWASDKGTYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPC 911

Qy        144 ESPPGS---------------SPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTC--- 184
Db        912 TLLAGNPIPERRWIKNSAMLLQNPYIITVRS--------DGSLHIERVQLQDGGEYTCVAS 963

Qy        185 TVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAER 244
Db        964 NVAGTNNKTTSVVVHVLPTIQHGQQILSTIEG--IPVTLP----------------CKA 1004

Qy        245 ASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK 304
Db       1005 SGNPKPSVIWSKKGELIS----TSSAKFSAGADGSLYVVSPEG----EESGEYVCTATNT 1056

Qy        305 TGKLHQEVNLVV---------MRATQLQK-----------NLTCEV---------W--- 331
Db       1057 AGYAKRKVQLTVYVRPRVFGDLRGLSQDKPVEISVLAGEEVTLPCEVKSLPPPIITWAKE 1116

Qy        332 ----GPTSPK--LMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSD-SGQV--LLES 382
Db       1117 TQLISPFSPRHTFLPSGSMKITETRTS----------DSGMYLCVATNIAGNVTQAVKL 1165

Qy        383 NIKVLPTWST-PVHPRASA-----LPAPPTGSALP 411
Db       1166 NVHVPPKIQRGPKHLKVQVGQRVDIPCNAQGTPLP 1200


RESULT 43
Q8K0X1
ID   Q8K0X1      PRELIMINARY;      PRT;   739 AA.
AC   Q8K0X1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vascular cell adhesion molecule 1.
GN   VCAM1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Strausberg R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC029823; AAH29823.1; -.
DR   MGD; MGI:98926; Vcam1.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0016337; P:cell-cell adhesion; IEA.
DR   InterPro; IPR003987; ICAM_VCAM-1.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003989; VCAM-1.
DR   Pfam; PF00047; ig; 5.
DR   PRINTS; PR01472; ICAMVCAM1.
DR   PRINTS; PR01474; VCAM1.
DR   SMART; SM00409; IG; 5.
DR   SMART; SM00408; IGc2; 5.
DR   PROSITE; PS50835; IG_LIKE; 5.
KW   Immunoglobulin domain.
SQ   SEQUENCE   739 AA;  81329 MW;  8B66F3AAF9DD720E CRC64;

  Query Match             6.1%;  Score 142;  DB 11;  Length 739;
  Best Local Similarity   20.2%;  Pred. No. 0.031;
  Matches   79;  Conservative   65;  Mismatches  147;  Indels  100;  Gaps   15;

Qy         32 KKGDTVELTCTASQKKSIQFHW--KNSNQI-KILGNQGSFLTKGPSKLNDRADSRRSLWD 88
Db        237 QEGGAVTMTCSSEGLPAPEIFWGRKLDNEVLQLL-------------------------- 270

Qy         89 QGNFPLIIKNLKIEDSDTYICE----VEDQKEEVQLLV------FGLTANSDTHLLQGQS 138
Db        271 SGNAILTLIAMRMEDSGVYVCEGVNLIGRDKAEVELVVQEKPFIVDISPGSQVAAQVGDS 330

Qy        139 LTLTLESPPGSSPSVQCRSPRGKNIQG-------GKTLSVSQLELQDSGTWTCTVLQNQK 191
Db        331 VVLTCAAIGCDSPSFSWRTQTDSPLNGVVRNEGAKSTLVLSSVGFEDEHSYLCAVTCLQR 390

Qy        192 KVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLA------------------FTVEKL 232
Db        391 TLEKRTQVEVYSF---------PEDPVIKMSGPLVHGRPVTVNCTVPNVYPFDHLEIELL 441

Qy        233 TGSGELWWQAERASSSKSWI-TFDLKNKEVSVKRVTQDPKLQ-MGKKL----PLHLTLPQ 286
Db        442 KG--------ETTLMKKYFLEEMGIKSLETKILETTFIPTIEDTGKSLVCLARLHSGEME 493

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQK----NLTCEVWGPTSPKLMLSL 342
Db        494 SEPKQRQSVQPLYVNVAP-----KETTIWVSPSPILEEGSPVNLTCSSDGIPAPKILWSR 548

Qy        343 KLENKEAKVSKREKPVWVLN---PEAGMWQC 370
Db        549 QLNNGELQPLSENTTLTFMSTKRDDSGIYVC 579


RESULT 44
Q8HYV1
ID   Q8HYV1      PRELIMINARY;      PRT;   885 AA.
AC   Q8HYV1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Glycosylphosphatidylinositol anchor 1 protein (Fragment).
GN   MDGA1.
```

```
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kierstein S., Al-Bayati H., Kollers S., Brenig B.;
RT   "Assignment of the porcine MAM domain containing
RT   glycosylphosphatidylinositol anchor 1(mdga1) on chromosome 7q11-13 by
RT   in stiu hybridisation.";
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ459296; CAD30702.2; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000998; MAM_domain.
DR   Pfam; PF00047; ig; 5.
DR   Pfam; PF00629; MAM; 1.
DR   PRINTS; PR00020; MAMDOMAIN.
DR   SMART; SM00409; IG; 5.
DR   SMART; SM00408; IGc2; 5.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS50835; IG_LIKE; 5.
DR   PROSITE; PS50060; MAM_2; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   885 AA;  97766 MW;  BC2929D59C3491C5 CRC64;

  Query Match             6.1%;  Score 142;  DB 6;  Length 885;
  Best Local Similarity   22.1%;  Pred. No. 0.039;
  Matches  107;  Conservative   68;  Mismatches  176;  Indels  134;  Gaps   25;

Qy         36 TVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI 95
              || | || :     :| ||  :             |   |: |   |     |: ||   ::
Db         82 TVFLRCTVNSNPPARFIWKRGSD-----------TLSHSQDNG-VDIYEPLYTQGGAEVL 129

Qy         96 -IKNLKIEDSDTYICEVED------QKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPG 148
               :|:|: ::   :  |:| |         : | | :   ||      |    |:  ||   ||
Db        130 KLKDLRPQEYASLTCQVSDGSVCGIPDKAVTLRLTSTTAPPALKL----SVNETLVVNPG 185

Qy        149 SSPSVQC-----------------RSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN- 189
              : :|||                 | |   ||| |||:  :: :||| : ||   |
Db        186 ENVTVQCLLQGGDPLTQLQWSHGPGPLPLGALAQGG-TLSIPSVQARDSGYYNCTATNNV 244

Qy        190 ---QKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFSFPL-------AFTVEKLTGSG 236
                 ||    :::: | : : |:  :     || | ::    |       |   ||:|
Db        245 GNPAKKT---VNLLVRSMKNATFQITPDVIKESENIQLGQDLKLSCHVDAVPQEKVTYQ- 300

Qy        237 ELWWQ-AERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHL----------TLP 285
               |::  : |  ||  :   :   :  :  ||   :|     : ||           : |
Db        301 --WFKNGKPARMSKRLL---VTRNDPELPAVTSSLEL-----IDLHFSDYGTYLCVASFP 350

Qy        286 QA-LPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLS--- 341
              | :|  :   |::      |  : :   :| :|      | ||| |   | :: |
Db        351 GAPVPDLSVEVNISSETVPPTISVPKGRAVVTVREGS-PAELQCEVRGKPRPPVLWSRVD 409

Qy        342 ---------LKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSG-------QVLLESNIK 385
                      | ||      : |:   |    :| ::|  :           :  :: |::
Db        410 KEAALLPSGLPLEETSDGKLRLER---VSRDMSGTYRCQTARYNGFNVRPREAQVQLNVQ 466

Qy        386 VLP-----------TWSTPVHPRASALPAPPTGSALPDPQTASA--------LPDPPAAS 426
               |                ||  | | |   |        : |||        || ||
Db        467 FPPEVEPGSQDVRQALGRPVLLRCSLLRGSP-------QRIASAVWRFKGQLLPPPP--- 516

Qy        427 ALPAA 431
              |:| |
Db        517 AIPIA 521


RESULT 45
Q8HYV2
ID   Q8HYV2      PRELIMINARY;      PRT;   886 AA.
AC   Q8HYV2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MAM domain containing glycosylphosphatidylinositol anchor 1
DE   (Fragment).
GN   MDGA1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kierstein S., Al-Bayati H., Kollers S., Brenig B.;
RT   "Assignment of the porcine MAM domain containing
RT   glycosylphosphatidylinositol anchor 1 (MDGA1) on chromosome 7q11-q13
RT   by in situ hybridisation.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ459295; CAD30701.2; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000998; MAM_domain.
DR   Pfam; PF00047; ig; 5.
DR   Pfam; PF00629; MAM; 1.
DR   PRINTS; PR00020; MAMDOMAIN.
DR   SMART; SM00409; IG; 5.
DR   SMART; SM00408; IGc2; 5.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS50835; IG_LIKE; 5.
DR   PROSITE; PS50060; MAM_2; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   886 AA;  98074 MW;  B669872711DB9D52 CRC64;

  Query Match             6.1%;  Score 142;  DB 6;  Length 886;
  Best Local Similarity   22.1%;  Pred. No. 0.04;
  Matches  107;  Conservative   68;  Mismatches  176;  Indels  134;  Gaps   25;

Qy         36 TVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI 95
              || | || :     :| ||  :             |   |: |   |     |: ||   ::
Db         83 TVFLRCTVNSNPPARFIWKRGSD-----------TLSHSQDNG-VDIYEPLYTQGGAEVL 130

Qy         96 -IKNLKIEDSDTYICEVED------QKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPG 148
               :|:|: ::   :  |:| |         : | | :   ||      |    |:  ||   ||
Db        131 KLKDLRPQEYASLTCQVSDGSVCGIPDKAVTLRLTSTTAPPALKL----SVNETLVVNPG 186

Qy        149 SSPSVQC-----------------RSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN- 189
              : :|||                 | |   ||| |||:  :: :||| : ||   |
Db        187 ENVTVQCLLQGGDPLTQLQWSHGPGPLPLGALAQGG-TLSIPSVQARDSGYYNCTATNNV 245

Qy        190 ---QKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFSFPL-------AFTVEKLTGSG 236
                 ||    :::: | : : |:  :     || | ::    |       |   ||:|
Db        246 GNPAKKT---VNLLVRSMKNATFQITPDVIKESENIQLGQDLKLSCHVDAVPQEKVTYQ- 301

Qy        237 ELWWQ-AERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHL----------TLP 285
               |::  : |  ||  :   :   :  :  ||   :|     : ||           : |
Db        302 --WFKNGKPARMSKRLL---VTRNDPELPAVTSSLEL-----IDLHFSDYGTYLCVASFP 351

Qy        286 QA-LPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLS--- 341
              | :|  :   |::      |  : :   :| :|      | ||| |   | :: |
Db        352 GAPVPDLSVEVNISSETVPPTISVPKGRAVVTVREGS-PAELQCEVRGKPRPPVLWSRVD 410

Qy        342 ---------LKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSG-------QVLLESNIK 385
                      | ||      : |:   |    :| ::|  :           :  :: |::
Db        411 KEAALLPSGLPLEETSDGKLRLER---VSRDMSGTYRCQTARYNGFNVRPREAQVQLNVQ 467

Qy        386 VLP-----------TWSTPVHPRASALPAPPTGSALPDPQTASA--------LPDPPAAS 426
               |                ||  | | |   |        : |||        || ||
Db        468 FPPEVEPGSQDVRQALGRPVLLRCSLLRGSP-------QRIASAVWRFKGQLLPPPP--- 517

Qy        427 ALPAA 431
```

```
                |:| |
Db        518 AIPIA 522


RESULT 46
Q9JIE0
ID   Q9JIE0      PRELIMINARY;      PRT;   397 AA.
AC   Q9JIE0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NK cell receptor 2B4.
GN   NMRK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A.CA;
RX   MEDLINE=20395298; PubMed=10941850;
RA   Kumaresan P.R., Huynh V.T., Mathew P.A.;
RT   "Polymorphism in the 2B4 gene of inbred mouse strains.";
RL   Immunogenetics 51:758-761(2000).
DR   EMBL; AF234831; AAF91290.1; -.
DR   MGD; MGI:109294; Nmrk.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   SMART; SM00409; IG; 1.
KW   Receptor.
SQ   SEQUENCE   397 AA;  44777 MW;  438A7432A50927E3 CRC64;

  Query Match             6.1%;  Score 141.5;  DB 11;  Length 397;
  Best Local Similarity   21.9%;  Pred. No. 0.014;
  Matches   85;  Conservative   61;  Mismatches  155;  Indels   87;  Gaps   14;

Qy         18 LLPAATQG------NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQ-----IKILGNQG 66
              ||  | ||      :: |:|  |  | |  :  | | :   || : |     |:||
Db         12 LLLRAHQGQDCPDSSEEVVGVSGKPVPLRPSNIQTKDVSVQWKKTEQGSHRKIEIL---- 67

Qy         67 SFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEV-----QLL 121
              ::   |||  |          :| |:| | ||: |::||  |: |: :   :|     |||
Db         68 NWYNDGPSWSNVSFSDIYG-FDYGDFALSIKSAKLQDSGHYLLEITNTGGKVCNKNFQLL 126

Qy        122 VFGLTANSDTHLLQGQ-------SLTLTLESPPGSSPSVQCRSPRGKNI---QGGKTLSV 171
              :    : :|  |: |       :  | |         :|      ||  :   |   |
Db        127 ILD---HVETPNLKAQWKPWTNGTCQLFLSCLATKDDNVSYALYRGSTLISNQRNSTHWE 183

Qy        172 SQLELQDSGTWTC-------------------TVLQNQKKVEFKIDIVVLAFQKASSI- 210
              :|::     |:||                 :|  | : : | : ||:|      :|
Db        184 NQIDASSLHTYTCNVSNRASWANHTLNFTHGCQSVPSNFRFLPFGVIIVILVTLFLGAII 243

Qy        211 ---VYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAE---RASSSKSWITFDLKNKEVSVK 264
                 |: |: :|::||     |: :           :   ||| | | :  | : :    :
Db        244 CFCVWTKKRKQLQFSPKEPLTIYEYVKDSRASRDQQGCSRASGSPSAVQEDGRGQRELDR 303

Qy        265 RVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQK 324
              ||::                : : |||     |:          | :: :       :| :|
Db        304 RVSE---------------VLEQLPQQTFPGD--------RGTMYSMIQCKPSDSTSQEK 340

Qy        325 NLTCEVWGPTSPKLMLSLKLENKEAKVS 352
                | |:    |      |  |: : :|
Db        341 ---CTVYSVVQPSRKSGSKKRNQNSSLS 365


RESULT 47
Q8N4D0
ID   Q8N4D0      PRELIMINARY;      PRT;   702 AA.
AC   Q8N4D0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Carcinoembryonic antigen-related cell adhesion molecule 5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon, and Kidney;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC034671; AAH34671.1; -.
DR   GO; GO:0003779; F:actin binding; IEA.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00409; IG; 7.
DR   SMART; SM00408; IGc2; 6.
DR   PROSITE; PS00019; ACTININ_1; 3.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   Immunoglobulin domain.
SQ   SEQUENCE   702 AA;  76781 MW;  97CCFB7399A0B05A CRC64;

  Query Match             6.0%;  Score 140.5;  DB 4;  Length 702;
  Best Local Similarity   19.0%;  Pred. No. 0.037;
  Matches   97;  Conservative   75;  Mismatches  215;  Indels  123;  Gaps   20;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
              |: :  |   :  | | |  ||    : :    | |: : :
Db        147 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV----------------- 187

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
                | |     ||  | : |:   || :| || ::     : : | | :::|  | :
Db        188 --SPRLQLSNGNRTLTLFNVTRNDSASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 245

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
              :|    |::| |:  :   |:|  |       |    |  : | :  : : :||::||
Db        246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 303

Qy        189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
              :  :  :  : :  :   :      :|    :| | | ::  :  |||   ::
Db        304 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 363

Qy        247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
                :     |: :::  ||::           :| :    |: |           |:  :
Db        364 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 423

Qy        288 LPQYAGSGNLTLALEAKT-----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
                |    ||:|:  | :           | : |    | :   |:    | ||:
Db        424 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 483

Qy        335 SPKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLE 381
              |     ::|     |::      |   |||       :   |||      |   :||
Db        484 SGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ---- 539

Qy        382 SNIKVLPTWSTPVHPRASALPAPPT---------------------GSALPDPQTASAL 419
                  | || ||         |                        :   || |   |
Db        540 ---------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVL 590

Qy        420 --PDPPAASALPAALAVISFLLGLGLGVAC 447
                || |  |   :     |:| |  | ::|
Db        591 YGPDTPIISPPDS-----SYLSGANLNLSC 615


RESULT 48
Q95YM2
ID   Q95YM2      PRELIMINARY;      PRT; 17352 AA.
AC   Q95YM2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   I-connectin.
```

```
GN   I-CON.
OS   Procambarus clarkii (Red swamp crayfish).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Astacoidea; Cambaridae; Procambarus.
OX   NCBI_TaxID=6728;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21423462; PubMed=11532946;
RA   Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA   Sun P., Maruyama K., Kimura S.;
RT   "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RT   sarcomeres of crayfish claw muscle.";
RL   EMBO J. 20:4826-4835(2001).
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; AB055861; BAB64297.1; -.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   InterPro; IPR000577; FGGY_kin.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 49.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00408; IGc2; 13.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00933; FGGY_KINASES_1; 3.
DR   PROSITE; PS50835; IG_LIKE; 49.
DR   PROSITE; PS50002; SH3; 1.
KW   Immunoglobulin domain; SH3 domain.
SQ   SEQUENCE   17352 AA;  1962348 MW;  4BA157BEC042E42D CRC64;

  Query Match             6.0%;  Score 140.5;  DB 5;  Length 17352;
  Best Local Similarity   21.1%;  Pred. No. 3.3;
  Matches   92;  Conservative   69;  Mismatches  156;  Indels  119;  Gaps   20;

Qy         37 VELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLII 96
              ||:|   ::| :: :| |:  :|      :   :::| ::                |::
Db      16247 VEVT---NEKATVTWH-KDGEEITEHHERFKLVSQGKTR-----------------KLVL 16285

Qy         97 KNLKIEDSDTYICEVEDQKEEVQLLVFGLTAN-----SDTHLLQGQSLTLTLESPPGSS- 150
                :|   | | : ||:   :| | | | |       |  : :|   |: :|   | :
Db      16286 MEATLSDEGEYTCVLGDQECTAELTVRELPAEIVRKMKDQVVSKGDRATMEVELTKGDAV 16345

Qy        151 -----PSVQCR-------SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKID 198
                  |: |       |  ||     : | |   : :||||:   | :::     |::
Db      16346 ITWYKDEVEIRFSDHYQLSIDGK----VQRLMVYNCQFEDSGTYRAVVGKSECSATLKVE 16401

Qy        199 IVVLAFQKASSIVYKKEGE---------QVEFSFPLAFTVEKLTGSGELWW---QAERAS 246
              : |              ||:         | |     | || :|   |: |    || :|
Db      16402 LQV-------------EGDFSKKLPAQMDVNFKTDATFVVE-ITKDYEVKWLREGAELSS 16447

Qy        247 SSKSWITFDLKNKEVSVKRVTQDPKLQ-------MGKKLPLHLTLPQALPQYAGSGNLTL 299
              | |  |  : | : : || |:|    :       :     ||:   :  |
Db      16448 SEKYVIKKEAKKRILIVKSVSQTDACEYSCVLGNLKTSCVLHVVRMETAP---------- 16497

Qy        300 ALEAKTGKLHQEVNLVVMRATQLQKNLTCEV--WGPTSPKLM-----LSLKLENKE---A 349
                  | ||:| ::| :     |:   :|      :||:         |  || |
Db      16498 ----KIPKEHQKVEVIVTKG----KDAVLKVPFTATPTPKVFWYHKGQLLNTENTEKLLP 16549

Qy        350 KVSKREKPV---WVLNPEAGMWQC-LLSDSGQVLLESNIKVLPTWSTPVHPRAS------ 399
               :|::|  :    | | : | ::  | :| |     :  :|:|   | |  |
Db      16550 TISEQEASITIKQVENIDCGEYRLKLCNDCGAAYADFTLKILDKPSQPGTPEPMEVTNVS 16609

Qy        400 -----ALPAPPTGSAL 410
                  :||    | |:
Db      16610 VTLHWSLPKEDGGRAI 16625


RESULT 49
Q9VP08
ID   Q9VP08      PRELIMINARY;      PRT;   403 AA.
AC   Q9VP08;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG7166 protein.
GN   CG7166.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
```

```
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003595; AAF51754.2; -.
DR   FlyBase; FBgn0037107; CG7166.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain.
SQ   SEQUENCE   403 AA;  45747 MW;  429C1039F539F76D CRC64;

  Query Match             6.0%;  Score 139.5;  DB 5;  Length 403;
  Best Local Similarity   24.4%;  Pred. No. 0.021;
  Matches   57;  Conservative   39;  Mismatches   97;  Indels   41;  Gaps   11;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||::   |:|:|| |       |    |:  :         | || |  |:    | |  :
Db         33 KVIV---GETIELPCKVQNLGSFVLLWRKGS---------SVLTAGHLKIT--RDQRFKI 78

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEE-----VQLLV----FGLTANSDTHLLQGQ 137
                 |:: | |  :| :|:  |||:: ||:       |::||      |  |      :|
Db         79 --VGDYNLQINGVKTQDAGDYICQLGDQENRDQVHTVEILVPPTLRALPHNGQVTARKGS 136

Qy        138 SLTLTLESPPGSSPSV------QCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQK 191
              ::||  ::     |::          |   ::    || :  ::   :||: |:     |
Db        137 TVTLECKASGNPVPTIFWFKKDVFSGP--THLSDSSTLILENVDRHHAGTYQCSADNGVK 194

Qy        192 -KVEFKIDIVVLA---FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQ 241
               :|   | : :|:        | |:  ||  ||    |   |     |  ||:|
Db        195 DRVSMDIQLTILSPPEITVEKSWVHASEGYDVE----LVCIVHGDVNSEMLWYQ 244


RESULT 50
Q07763
ID   Q07763      PRELIMINARY;      PRT;   398 AA.
AC   Q07763;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   2B4 protein precursor.
GN   NMRK OR 2B4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=94044757; PubMed=8228228;
RA   Mathew P.A., Garni-Wagner B.A., Land K., Takashima A., Stoneman E.,
RA   Bennett M., Kumar V.;
RT   "Cloning and characterization of the 2B4 gene encoding a molecule
RT   associated with non-MHC-restricted killing mediated by activated
RT   natural killer cells and T cells.";
RL   J. Immunol. 151:5328-5337(1993).
DR   EMBL; L19057; AAA16353.1; -.
DR   PIR; I49443; I49443.
DR   MGD; MGI:109294; Nmrk.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   SMART; SM00409; IG; 1.
KW   Signal.
FT   SIGNAL       1     18       POTENTIAL.
SQ   SEQUENCE   398 AA;  45132 MW;  5C6CDDACAE33525C CRC64;

  Query Match             5.9%;  Score 138;  DB 11;  Length 398;
  Best Local Similarity   21.0%;  Pred. No. 0.026;
  Matches   83;  Conservative   64;  Mismatches  147;  Indels  102;  Gaps   15;

Qy         18 LLPAATQG------NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQ-----IKILGNQG 66
              ||  | ||      :: |:|  |  |:|  :  | | :   || : |     |:||
Db         12 LLLRAHQGQDCPDSSEEVVGVSGKPVQLRPSNIQTKDVSVQWKKTEQGSHRKIEIL---- 67

Qy         67 SFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEV-----QLL 121
              ::   |||  |          :| |:| | ||: |::||  |: |: :   :|     |||
Db         68 NWYNDGPSWSNVSFSDIYG-FDYGDFALSIKSAKLQDSGHYLLEITNTGGKVCNKNFQLL 126

Qy        122 VFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQ----CRSPRGKNI-------------- 163
              :      |:   ::  |  :  | :: : |    |   :  |:
Db        127 IL-------DHV---ETPNLKAQWKPWTNGTCQLFLSCLVTKDDNVSYAFWYRGSTLISN 176

Qy        164 QGGKTLSVSQLELQDSGTWTC--------------------TVLQNQKKVEFKIDIVVLA 203
              |   |   :|::     |:||                    :|  | : : | : ||:|
Db        177 QRNSTHWENQIDASSLHTYTCNVSNRASWANHTLNFTHGCQSVPSNFRFLPFGVIIVILV 236

Qy        204 FQKASSI----VYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAE---RASSSKSWITFDL 256
                 :|    |: |: :|::||     |: :           :   ||| | | :  |
Db        237 TLFLGAIICFCVWTKKRKQLQFSPKEPLTIYEYVKDSRASRDQQGCSRASGSPSAVQEDG 296

Qy        257 KNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVV 316
              : :   :||::              : : |||    |:          | ::  :
Db        297 RGQRELDRRVSE--------------VLEQLPQQTFPGD--------RGTMYSMIQCKP 333

Qy        317 MRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVS 352
               :|  :|   | |:    |      |  |:   :|
Db        334 SDSTSQEK---CTVYSVVQPSRKSGSKKRNQNYSLS 366


RESULT 51
Q9NR99
ID   Q9NR99      PRELIMINARY;      PRT;  2828 AA.
AC   Q9NR99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Adlican.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Crowl R.M., Luk D.;
RT   "Identification of the gene encoding Adlican, a novel protein
RT   expressed in human arthritic tissues.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF245505; AAF86402.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR_typ.
```

```
DR   Pfam; PF00047; ig; 12.
DR   Pfam; PF00560; LRR; 6.
DR   Pfam; PF01463; LRRCT; 1.
DR   SMART; SM00408; IGc2; 10.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 3.
DR   PROSITE; PS50835; IG_LIKE; 12.
KW   Immunoglobulin domain.
SQ   SEQUENCE   2828 AA;  312291 MW;  A18377D8554F1FE1 CRC64;

  Query Match             5.9%;  Score 138;  DB 4;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.4;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
              | : | :| ||       |:  :    | :: : |||          |       :||
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
                  |    | | :  ||   ||   | |:||  :||  | |      |   :
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
              ||| |    ||:        | :  | :| |   :      | :  | |  : |
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
                    || |  :  :|:| : | | :|:      |  |:|:|:    | : | :|:
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE--------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
              | :                    :: :| |    |     |       :   :| ::
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
              |:   |   || ::                        |: |   | : ||
Db       2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT--------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
               | : |: |    :| |                 ||  |   ||:   |   ||    |  :
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
               ::             ::| : ||: :|       | :  |::
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ--------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
               ||  :  |:| |                 |  :| |    | |  : :
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 52
Q9QY38
ID   Q9QY38      PRELIMINARY;      PRT;  1259 AA.
AC   Q9QY38;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neural cell adhesion molecule L1.
GN   L1CAM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Platzer M., Brenner V., Reichwald K., Wiehe T., Oksche A.,
RA   Rosenthal A.;
RT   "Comparative sequence analysis of the mouse L1cam locus and the
RT   corresponding region of human Xq28.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF133093; AAF22153.1; -.
DR   HSSP; P20241; 1CFB.
DR   MGD; MGI:96721; L1cam.
DR   GO; GO:0007411; P:axon guidance; IMP.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 4.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00060; FN3; 3.
DR   SMART; SM00408; IGc2; 5.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   Immunoglobulin domain.
SQ   SEQUENCE   1259 AA;  140916 MW;  25743C039892A22F CRC64;

  Query Match             5.9%;  Score 137.5;  DB 11;  Length 1259;
  Best Local Similarity   20.0%;  Pred. No. 0.14;
  Matches   73;  Conservative   61;  Mismatches  136;  Indels   95;  Gaps   16;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |:|  | |    :   :  |:                  | |      | :  : :||:
Db       346 GETARLDCQVQGRPQPEITWR--------------INGMSMETVNKDQKYRI-EQGS-- 387

Qy         94 LIIKNLKIEDSDTYICEVEDQKEE----------VQLLVFGLTANSDTHL-LQGQSLTL 141
              ||: |  :: ||| : : | : :             |||    || :: |:: ::| :  |
Db       388 LILSN--VQPSDTMVTQCEARNQHGLLLANAYIYVVQLPARILTKDNQTYMAVEGSTAYL 445

Qy        142 TLESPPGSSPSVQCRSPRGKNI--------QGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
               ::     ||||     |  :            |||:  |:  |:| : |    :|  |
Db       446 LCKAFGAPVPSVQWLDEEGTTVLQDERFFPYANGTLSIRDLQANDTGRYFCQAANDQNNV 505

Qy        194 EFKIDIVVLAFQKASSI------VYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASS 247
                 :: |   ::|: |        :|:| :| |:   :|          : |: :
Db       506 TILANLQV---KEATQITQGPRSAIEKKGARVTFTCQASFDPSL---QASITWRGDGR-- 557

Qy        248 SKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGK 307
                   ||: :  | |   :| || :                |:|  |:  || :     :  :
Db       558 -------DLQERGDSDKYFIEDGKLVI----------QSL-DYSDQGNYSCVASTELDE 598

Qy        308 LHQEVNLVVMRA--------------TQLQKNLTCEVWGP----TSPKLMLSLKLENKE 348
              :     |:|: :                | | :|:   | |     ||     :: |:||
Db       599 VESRAQLLVVGSPGPVPHLELSDRHLLKQSQVHLS---WSPAEDHNSPIEKYDIEFEDKE 655

Qy        349 AKVSK 353
                |
Db       656 MAPEK 660


RESULT 53
O44924
ID   O44924      PRELIMINARY;      PRT;  1395 AA.
AC   O44924;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ROUNDABOUT 1.
GN   ROBO OR ROBO1 OR CG13521.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98117249; PubMed=9458045;
RA   Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,
RA   Goodman C.S., Tear G.;
RT   "Roundabout controls axon crossing of the CNS midline and defines a
RT   novel subfamily of evolutionarily conserved guidance receptors.";
RL   Cell 92:205-215(1998).
DR   EMBL; AF040989; AAC38849.1; -.
DR   HSSP; P56276; 1TLK.
DR   FlyBase; FBgn0005631; robo.
```

```
DR   GO; GO:0005886; C:plasma membrane; IDA.
DR   GO; GO:0007411; P:axon guidance; IMP.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF00047; ig; 5.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 3.
DR   SMART; SM00408; IGc2; 5.
DR   PROSITE; PS50835; IG_LIKE; 5.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1395 AA;  151778 MW;  B820E234A5218983 CRC64;

  Query Match             5.9%;  Score 136.5;  DB 5;  Length 1395;
  Best Local Similarity   21.1%;  Pred. No. 0.2;
  Matches  102;  Conservative   53;  Mismatches  179;  Indels  149;  Gaps   18;

Qy         26 NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS 85
              || | |  |  |:| | ||       |                :| | |     |
Db        353 NKKV-GLNG-VVQLPCMASGNPPPSVFWTK--------------EGVSTLMFPNSSHGR 395

Qy         86 LWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTH---LLQGQSLTLT 142
               :   : | | ::: ||   |:|      :   : ||   :: |     ::|       |
Db        396 QYVAADGTLQITDVRQEDEGYYVCSAFSVVDSSTVRVFLQVSSVDERPPPIIQIGPANQT 455

Qy        143 LESPPGSSPSVQCR-----SPRGK-------------NIQGGKTLSVSQLELQDSGTWT 183
              |  | ||  :: ||     |||| |              :|  | :| |  |:| ||||:|
Db        456 L--PKGSVATLPCRATGNPSPRIKWFHDGHAVQAGNRYSIIQGSSLRVDDLQLSDSGTYT 513

Qy        184 CTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAE 243
              ||                            ||: | |:    ||||  ||  |   |:
Db        514 CTA--------------------------SGERGETSWAATLTVEK-PGSTSLHRAAD 544

Qy        244 ----------------RASSSKSWITFDLKNKEVS-VKRVTQD---PKLQMGKKLPLH- 281
                             | | |  |     |   |  :   | :   | ||| |  :  |
Db        545 PSTYPAPPGTPKVLNVSRTSISLRWAKSQEKPGAVGPIIGYTVEYFSPDLQTGWIVAAHR 604

Qy        282 ---------------------------LTLPQALPQY----------AGSGNLTLALEA 303
                                        :::|  |             | : :|: |
Db        605 VGDTQVTISGLTPGTSYVFLVRAENTQGISVPSGLSNVIKTIEADFDAASANDLSAARTL 664

Qy        304 KTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNP 363
               |||  : ::   : |: ::      |      | :  |::  |:| |           |
Db        665 LTGKSVELIDASAINASAVRLEWMLHV--SADEKYVEGLRIHYKDASV----------P 711

Qy        364 EAGMWQCLLSD-SGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDP 422
               |      : | |  :  :|:|    :   :  |    :  |: |     :||    |
Db        712 SAQYHSITVMDASAESFVVGNLKKYTKYEFFLTPFFETIEGQPSNS-----KTALTYEDV 766

Qy        423 PAA 425
              |:|
Db        767 PSA 769


RESULT 54
Q7TQG5
ID   Q7TQG5      PRELIMINARY;      PRT;  1465 AA.
AC   Q7TQG5;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC054540; AAH54540.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   1465 AA;  159967 MW;  AB51C378BC3FBA32 CRC64;

  Query Match             5.9%;  Score 136.5;  DB 11;  Length 1465;
  Best Local Similarity   22.8%;  Pred. No. 0.21;
  Matches  115;  Conservative   53;  Mismatches  191;  Indels  145;  Gaps   27;

Qy         10 LLLVLQLALL---------------PAATQGNKV--------------VLGKKGDTVELT 40
              |||:| | ||                |  : |  |              |  :| :| |
Db         25 LLLLLPLLLLLGRPASGAAATKSGSPPQSAGASVRTFTPFYFLVEPVDTLSVRGSSVILN 84

Qy         41 CTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNL- 99
              |:|  : |    ||           |:|        ||  :| || |   |:  | | |:
Db         85 CSAYSEPSPNIEWK---------KDGTF-------LNLESDDRRQLLPDGS--LFISNVV 126

Qy        100 ----KIEDSDTYIC--EVED----QKEEVQLLVFGL----TANSDTHLLQGQSLTLTLES 145
                  |   | |   |::        :| | ||    :    : :  | |  |  |
Db        127 HSKHNKPDEGFYQCVATVDNLGTIVSRTAKLTVAGLPRFTSQPEPSSVYVGNSAILNCEV 186

Qy        146 PPGSSPSV---QCRSP-----RGKNIQGGKTLSVSQLELQDSGTWTCTVLQN-----QKK 192
                 | |   | | |      |   :  | || :|      | | : | |          :
Db        187 NADLVPFVRWEQNRQPLLLDDRIVKLPSG-TLVISNATEGDGGLYRCIVESGGPPKFSDE 245

Qy        193 VEFKI-----DIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASS 247
               | |:     :|| | |     | : |  |:    |  |   |  |     |
Db        246 AELKVLQDPEEIVDLVFLMRPSSMMKVTGQ----SAVLPCVVSGLPAPVVRW-------- 293

Qy        248 SKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQY---AGSGNLT------ 298
                    :||:||     :    |  | :| :           |   | :|| |
Db        294 --------MKNEEVLDTESSGRLVLLAGGCLEISDVTEDDAGTYFCIADNGNKTVEAQAE 345

Qy        299 LALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLML----SLKLENKEAKVSKR 354
              | ::   | | |  |:     :  :     ||| |  :| :        :  :   |: |
Db        346 LTVQVPPGFLKQPANIYAHESMDIV--FECEVTGKPTPTVKWVKNGDVVIPSDYFKIVKE 403

Qy        355 E--KPVWVLNPEAGMWQCLL-SDSGQ-------VLLESNIKVLPTWSTPVHPRASALPAP 404
              : : ::  : | :||:  :| |          ::|| ::  :||              |
Db        404 HNLQVLGLVKSDEGFYQCIAENDVGNAQAGAQLIILEHDV-AIPT-------------LP 449

Qy        405 PTG--SALPD---PQTASALPDPP 423
              ||   ||  |   | |   ||  |
Db        450 PTSLTSATTDHLAPATTGPLPSAP 473


RESULT 55
Q8IRV9
ID   Q8IRV9      PRELIMINARY;      PRT;  4117 AA.
```

```
AC   Q8IRV9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG7981-PC.
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003424; AAN09077.1; -.
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006306; P:DNA methylation; IEA.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00047; ig; 12.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 23.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 9.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00409; IG; 12.
DR   SMART; SM00408; IGc2; 12.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 23.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50835; IG_LIKE; 11.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 20.
DR   PROSITE; PS50068; LDLRA_2; 23.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
SQ   SEQUENCE   4117 AA;  454841 MW;  BB47056B7E4D6106 CRC64;

  Query Match             5.9%;  Score 136.5;  DB 5;  Length 4117;
  Best Local Similarity   17.8%;  Pred. No. 0.89;
  Matches   61;  Conservative   61;  Mismatches  109;  Indels  111;  Gaps   13;

Qy         19 LPAATQGNKVVLG------KKGDTVELTCTASQKKSIQFHWKN------SNQIKILGNQG 66
              ||| ::   : |       : |:: :: | :|       |:       || :: :||:
Db       3229 LPAKSRDYSLKLDDQSSNLRAGESTDVECYSSDDTYTDVVWERSDGAPLSNNVRQVGNR- 3287

Qy         67 SFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLT 126
                                          |:| |:   |:  |:|: :  : ::   : |
Db       3288 --------------------------LVISNVSPSDAGNYVCKCKTDEGDLYTTSYKLE 3320

Qy        127 ANSDTHLLQGQSLTLTLESPPGSSPSVQC-----RSPRGK------NIQGGKT-----LS 170
                 | |:   :   : :  |::  :||     | |  :       :| |::     ||
Db       3321 VEDQPHELKSSKI---VYAKVGANADLQCGADESRQPTYRWSRQYGQLQAGRSLMNEKLS 3377
```

```
Qy        171 VSQLELQDSGTWTCTV-LQNQKKVEFKIDIVV-------------LAFQKASSIVYKKE 215
                :  ::  |:||: ||    : :  :|   :||              ::|    :  :|
Db       3378 LDSVQANDAGTYICTAQYADGETADFPNILVVTGAIPQFRQEPRSYMSFPTLPNSSFK-- 3435

Qy        216 GEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNK--------------- 259
                |:| | |  |   | | | :  :   |   :|   ||::
Db       3436 -----FNFELTFRPE--NGDGLLLFNGQTRGSG-DYIALSLKDRYAEFRFDFGGKPMLVR 3487

Qy        260 ----------EVSVKRVTQDPKLQMGKKLPLHLTLPQALPQ 290
                       | | |  :|  :|: :: |:     | :||
Db       3488 AEEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQ 3529


RESULT 56
Q9W4Y4
ID   Q9W4Y4      PRELIMINARY;      PRT;  4179 AA.
AC   Q9W4Y4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG7981 protein.
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003424; AAF45786.3; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006306; P:DNA methylation; IEA.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00047; ig; 12.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 23.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   SMART; SM00180; EGF_Lam; 9.
DR   SMART; SM00409; IG; 12.
DR   SMART; SM00408; IGc2; 12.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 23.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50835; IG_LIKE; 11.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 20.
DR   PROSITE; PS50068; LDLRA_2; 23.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Immunoglobulin domain; Laminin EGF-like domain.
```

```
SQ   SEQUENCE   4179 AA;  461782 MW;  47804277D0914E63 CRC64;

  Query Match             5.9%;  Score 136.5;  DB 5;  Length 4179;
  Best Local Similarity   17.8%;  Pred. No. 0.91;
  Matches   61;  Conservative   61;  Mismatches  109;  Indels  111;  Gaps   13;

Qy         19 LPAATQGNKVVLG------KKGDTVELTCTASQKKSIQFHWKN------SNQIKILGNQG 66
              ||| ::   : |       : |:: :: | :|        |:       || :: :||:
Db       3291 LPAKSRDYSLKLDDQSSNLRAGESTDVECYSSDDTYTDVVWERSDGAPLSNNVRQVGNR- 3349

Qy         67 SFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLT 126
                                      |:| |:   |:  |:|: :  : ::    : |
Db       3350 --------------------------LVISNVSPSDAGNYVCKCKTDEGDLYTTSYKLE 3382

Qy        127 ANSDTHLLQGQSLTLTLESPPGSSPSVQC-----RSPRGK------NIQGGKT-----LS 170
               | |:   :   : :  |::  :||     | |  :       :| |::     ||
Db       3383 VEDQPHELKSSKI---VYAKVGANADLQCGADESRQPTYRWSRQYGQLQAGRSLMNEKLS 3439

Qy        171 VSQLELQDSGTWTCTV-LQNQKKVEFKIDIVV--------------LAFQKASSIVYKKE 215
              :  ::  |:||: ||    : :  :|   :||             ::|    :  :|
Db       3440 LDSVQANDAGTYICTAQYADGETADFPNILVVTGAIPQFRQEPRSYMSFPTLPNSSFK-- 3497

Qy        216 GEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNK--------------- 259
                |:| | |   |   | | | :  :   |   :|   ||::
Db       3498 -----FNFELTFRPE--NGDGLLLFNGQTRGSG-DYIALSLKDRYAEFRFDFGGKPMLVR 3549

Qy        260 -----------EVSVKRVTQDPKLQMGKKLPLHLTLPQALPQ 290
                        | | |  :|  :|: :: |:     | :||
Db       3550 AEEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQ 3591


RESULT 57
Q8MPN3
ID   Q8MPN3      PRELIMINARY;      PRT;  4223 AA.
AC   Q8MPN3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Perlecan (CG7981-PD).
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Voigt A.;
RT   "Perlecan participates in proliferation activation of quiescent
RT   Drosophila neuroblasts.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ487018; CAD31650.1; -.
DR   EMBL; AE003424; AAN09079.1; -.
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006306; P:DNA methylation; IEA.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000504; RNA_rec_mot.
```

```
DR   Pfam; PF00047; ig; 13.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 23.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00180; EGF_Lam; 8.
DR   SMART; SM00409; IG; 13.
DR   SMART; SM00408; IGc2; 13.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 23.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50835; IG_LIKE; 12.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 20.
DR   PROSITE; PS50068; LDLRA_2; 23.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   4223 AA;  466919 MW;  31D7C2B5C9B1D2E0 CRC64;

  Query Match             5.9%;  Score 136.5;  DB 5;  Length 4223;
  Best Local Similarity   17.8%;  Pred. No. 0.92;
  Matches   61;  Conservative   61;  Mismatches  109;  Indels  111;  Gaps   13;

Qy         19 LPAATQGNKVVLG------KKGDTVELTCTASQKKSIQFHWKN------SNQIKILGNQG 66
              ||| ::   : |       : |:: :: | :|         |:       || :: :||:
Db       3335 LPAKSRDYSLKLDDQSSNLRAGESTDVECYSSDDTYTDVVWERSDGAPLSNNVRQVGNR- 3393

Qy         67 SFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLT 126
                                          |:| |:   |:  |:|: :  : ::    : |
Db       3394 ---------------------------LVISNVSPSDAGNYVCKCKTDEGDLYTTSYKLE 3426

Qy        127 ANSDTHLLQGQSLTLTLESPPGSSPSVQC-----RSPRGK------NIQGGKT-----LS 170
                | |:  :   :  :  |::  :||      | |  :       :| |::      ||
Db       3427 VEDQPHELKSSKI---VYAKVGANADLQCGADESRQPTYRWSRQYGQLQAGRSLMNEKLS 3483

Qy        171 VSQLELQDSGTWTCTV-LQNQKKVEFKIDIVV--------------LAFQKASSIVYKKE 215
              :  ::  |:||: ||    : :  :|   :||               ::|    :  :|
Db       3484 LDSVQANDAGTYICTAQYADGETADFPNILVVTGAIPQFRQEPRSYMSFPTLPNSSFK-- 3541

Qy        216 GEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNK---------------- 259
                  |:| | |   | | | :  :   |   :|   ||::
Db       3542 -----FNFELTFRPE--NGDGLLLFNGQTRGSG-DYIALSLKDRYAEFRFDFGGKPMLVR 3593

Qy        260 -----------EVSVKRVTQDPKLQMGKKLPLHLTLPQALPQ 290
                        | | |  :|  :|: :: |:      | :||
Db       3594 AEEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQ 3635


RESULT 58
Q8IRV8
ID   Q8IRV8      PRELIMINARY;      PRT;  4228 AA.
AC   Q8IRV8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG7981-PE.
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003424; AAN09078.1; -.
```

```
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0006306; P:DNA methylation; IEA.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00047; ig; 12.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 24.
DR   PRINTS; PR00261; LDLRECEPTOR.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF; 9.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00409; IG; 12.
DR   SMART; SM00408; IGc2; 12.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 24.
DR   PROSITE; PS00094; C5_MTASE_1; 1.
DR   PROSITE; PS00022; EGF_1; 10.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50835; IG_LIKE; 11.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 21.
DR   PROSITE; PS50068; LDLRA_2; 24.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
SQ   SEQUENCE   4228 AA;  466934 MW;  0F630AA0BBF4CD33 CRC64;

  Query Match             5.9%;  Score 136.5;  DB 5;  Length 4228;
  Best Local Similarity   17.8%;  Pred. No. 0.93;
  Matches   61;  Conservative   61;  Mismatches  109;  Indels  111;  Gaps   13;

Qy         19 LPAATQGNKVVLG------KKGDTVELTCTASQKKSIQFHWKN------SNQIKILGNQG 66
              ||| ::   : |       : |:: :: | :|        |:       || :: :||:
Db       3340 LPAKSRDYSLKLDDQSSNLRAGESTDVECYSSDDTYTDVVWERSDGAPLSNNVRQVGNR- 3398

Qy         67 SFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLT 126
                                          |:| |:   |:  |:|: :  : ::    : |
Db       3399 --------------------------LVISNVSPSDAGNYVCKCKTDEGDLYTTSYKLE 3431

Qy        127 ANSDTHLLQGQSLTLTLESPPGSSPSVQC-----RSPRGK------NIQGGKT-----LS 170
               | |:   :   : :  |::  :||     | |  :       :| |::     ||
Db       3432 VEDQPHELKSSKI---VYAKVGANADLQCGADESRQPTYRWSRQYGQLQAGRSLMNEKLS 3488

Qy        171 VSQLELQDSGTWTCTV-LQNQKKVEFKIDIVV--------------LAFQKASSIVYKKE 215
              :  ::  |:||: ||    : :  :|   :||               ::|    :  :|
Db       3489 LDSVQANDAGTYICTAQYADGETADFPNILVVTGAIPQFRQEPRSYMSFPTLPNSSFK-- 3546

Qy        216 GEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNK--------------- 259
                |:| | |  |   | | | :  :   |   :|   ||::
Db       3547 -----FNFELTFRPE--NGDGLLLFNGQTRGSG-DYIALSLKDRYAEFRFDFGGKPMLVR 3598

Qy        260 -----------EVSVKRVTQDPKLQMGKKLPLHLTLPQALPQ 290
                        | | |  :|  :|: :: |:     | :||
Db       3599 AEEPLALNEWHTVRVSRFKRDGYIQVDEQHPVAFPTLQQIPQ 3640


RESULT 59
Q96DN8
ID   Q96DN8      PRELIMINARY;      PRT;   512 AA.
AC   Q96DN8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ31774.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA   Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA   Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA   Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA   Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA   Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK056336; BAB71154.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00408; IGc2; 5.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   Hypothetical protein; Immunoglobulin domain.
SQ   SEQUENCE   512 AA;  54971 MW;  9368150E8A5BD6C8 CRC64;

  Query Match             5.8%;  Score 136;  DB 4;  Length 512;
  Best Local Similarity   20.9%;  Pred. No. 0.053;
  Matches   86;  Conservative   48;  Mismatches  150;  Indels  128;  Gaps   17;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              | | | |           |:  : : |      |     |:|   |
Db        148 GSNVTLPCYVQGYPEPTIKWRRLDNMPIFSR--PFSVSSISQLRTGA------------- 192

Qy         94 LIIKNLKIEDSDTYICEVEDQ----KEEVQLLVFGLTA------NSDTHLLQGQSLTLTL 143
              | | ||   |  |||| |:|    : |  : | || |       |  ::::|| |||
Db        193 LFILNLWASDKGTYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPC 252

Qy        144 ESPPGS---------------SPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTC--- 184
                 |:                :| :  ||          :| : :::||| | :||
Db        253 TLLAGNPIPERRWIKNSAMLLQNPYITVRS--------DGSLHIERVQLQDGGEYTCVAS 304

Qy        185 TVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAER 244
              |   |    :  |:  |    |:   || :  : |                      :
Db        305 NVAGTNNKTTSVVVHVLPTIQHGQQILSTIEG--IPVTLP-----------------CKA 345

Qy        245 ASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK 304
              : : |  : :  | : :|    |   |   |    |::  |        ||
Db        346 SGNPKPSVIWSKKGELIS----TSSAKFSAGADGSLYVVSPGG----EESGEYVCTATNT 397

Qy        305 TGKLHQEVNLVVM---RATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVL 361
              |   ::| | |     |   |: |:                   ::|  ::|       ||
Db        398 AGYAKRKVQLTVYVRPRVFGDQRGLS----------------QDKPVEIS-------VL 433

Qy        362 NPEAGMWQCLLSDSGQVLLESNIKVLP----TWS------TPVHPRASALPA 403
              |            :| |   :| ||    ||:      :|  || : ||:
Db        434 AGE-----------EVTLPCEVKSLPPPIITWAKETQLISPFSPRHTFLPS 473


RESULT 60
Q9VCT4
ID   Q9VCT4      PRELIMINARY;      PRT;   545 AA.
AC   Q9VCT4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   KLG protein (LD10776P).
GN   KLG OR CG6669.
```

```
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003741; AAF56071.1; -.
DR   EMBL; AY060363; AAL25402.1; -.
DR   FlyBase; FBgn0017590; klg.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain.
SQ   SEQUENCE   545 AA;  60087 MW;  D5E81D9E5574E9DE CRC64;

  Query Match             5.8%;  Score 136;  DB 5;  Length 545;
  Best Local Similarity   20.4%;  Pred. No. 0.058;
  Matches  105;  Conservative   84;  Mismatches  208;  Indels  118;  Gaps   23;

Qy          2 NRGVPFRHLLLVLQLAL--------LPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHW 53
              |||   | :  | | |:         ||             |||: | |      :   |
Db         75 NRGSNSRSMSNVQQSAVAASTLTATLPRFLSRGHTYRAVVGDTLVLPCQVENLGNFVLLW 134

Qy         54 KNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEV-- 111
              :    :    |    :|:          | |  | |   : | | :|: :|:  |:|::
Db        135 RRGTNVLTASN--IMVTR---------DERVRLID--GYNLEISDLEPQDAGDYVCQISD 181

Qy        112 ---EDQKEEVQLL----VFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKN-- 162
                 ||   |::|    |  :  :       :|  :||  :      ||:       | |
Db        182 KINRDQVHTVEILVPPSVRAIPTSGQLQARKGGPITLECKGSGNPVPSIYWTKKSGANKS 241

Qy        163 ---IQGGKTLSVSQLELQDSGTWTCTV---LQNQKKVEFKIDIVVLA-FQKASSIVYKKE 215
                 |  |  |:: :|| | :| : ||    : :   |: ::|::     | | ::  |
Db        242 TARIGDGPILTLEKLERQQAGVYQCTADNGVGDPVTVDMRLDVLYPPDIQVEKSWIHSGE 301

Qy        216 GEQVEF----------------SFPLAFTVEKLTGSGELWWQAERASSSKSWITF-DLK 257
              | : :                |||:  |  ::         | ||  ::  :|   ::
Db        302 GFEAKLVCIVFADPVATVSWYQNSFPIQSTDRRI--------MATRA--NRHMLTIRHIQ 351

Qy        258 NKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYA-------GSGNLTLALEAKTGKLHQ 310
               ::        | |   :|       | |   |:        | |||  :::      :
Db        352 QEDFGNYSCVADNSLGRSRKYMELSGRPGAAEFYSPKWGRSPDSYNLTWKIDSYPPL--E 409

Qy        311 EVNLVVMRATQLQKNLTCEVWG-----------PTSPKL--MLSLKLEN------KEAK 350
              || |:  |   :| | | :  |           | |  |  ::|  ::|        ||
Db        410 EVRLLYRR---VQMNETYQQPGRWHDFILTPEHRPASEPLTHIMSYTIKNLHPGGYYEAI 466

Qy        351 VSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSAL 410
              |  : :  |  |  : ::| ::: : | :   : :|:   |:     :::: : |  : |
Db        467 VQAKNRYGW--NEVSDIFQFVVATNSQDIGPEDAEVVA--SSKSRSNSASIGSLPGSTVL 522

Qy        411 PDPQTASALPDPPAASALPAALAVISF---LLGLG 442
                          |  |  ||:   |    ||||
Db        523 ------------HTALLLALALSNCQFDRRRLGLG 545


RESULT 61
Q9NEG1
ID   Q9NEG1      PRELIMINARY;      PRT;  1035 AA.
AC   Q9NEG1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   EG:BACR25B3.11 protein.
GN   TROL OR EG:BACR25B3.11 OR CG7981.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Murphy L., Harris D., Barrell B.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL138972; CAB72284.1; -.
DR   HSSP; P00740; 1EDM.
DR   FlyBase; FBgn0001402; trol.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR002172; LDL_receptor_A.
```

```
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00047; ig; 2.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 1.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00192; LDLa; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS50835; IG_LIKE; 2.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 1.
DR   PROSITE; PS50068; LDLRA_2; 1.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   EGF-like domain; Immunoglobulin domain.
SQ   SEQUENCE   1035 AA;  114044 MW;  413DE7EDD3AEE711 CRC64;

  Query Match             5.8%;  Score 134.5;  DB 5;  Length 1035;
  Best Local Similarity   17.6%;  Pred. No. 0.19;
  Matches   57;  Conservative   58;  Mismatches  103;  Indels  105;  Gaps   12;

Qy         32 KKGDTVELTCTASQKKSIQFHWKN------SNQIKILGNQGSFLTKGPSKLNDRADSRRS 85
              : |:: :: | :|       |:       || :: :||:
Db        179 RAGESTDVECYSSDDTYTDVVWERSDGAPLSNNVRQVGNR-------------------- 218

Qy         86 LWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLES 145
                    |:| |:   |:  |:|: :  : ::    : |      | |:   :   : :
Db        219 --------LVISNVSPSDAGNYVCKCKTDEGDLYTTSYKLEVEDQPHELKSSKI---VYA 267

Qy        146 PPGSSPSVQC-----RSPRGK------NIQGGKT-----LSVSQLELQDSGTWTCTV-LQ 188
                |::  :||     | |  :       :| |::     ||:  ::  |:||: ||
Db        268 KVGANADLQCGADESRQPTYRWSRQYGQLQAGRSLMNEKLSLDSVQANDAGTYICTAQYA 327

Qy        189 NQKKVEFKIDIVV-------------LAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTG 234
              : :  :|   :||             ::|    :  :|       |:| | | |   |
Db        328 DGETADFPNILVVTGAIPQFRQEPRSYMSFPTLPNSSFK-------FNFELTFRPE--NG 378

Qy        235 SGELWWQAERASSSKSWITFDLKNK-------------------------EVSVKRVT 267
                | | :  :   |   :|   ||::                         | | |
Db        379 DGLLLFNGQTRGSG-DYIALSLKDRYAEFRFDFGGKPMLVRAEEPLALNEWHTVRVSRFK 437

Qy        268 QDPKLQMGKKLPLHLTLPQALPQ 290
              :|  :|: :: |:     | :||
Db        438 RDGYIQVDEQHPVAFPTLQQIPQ 460


RESULT 62
Q7T2H2
ID   Q7T2H2      PRELIMINARY;      PRT;   487 AA.
AC   Q7T2H2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Fibroblast growth factor receptor-like protein precursor.
GN   FGFRL1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cartilage;
RA   Trueb B., Zhuang L., Taeschler S., Wiedemann M.;
RT   "Characterization of FGFRL1, a Novel FGF receptor preferentially
RT   expressed in skeletal tissues.";
RL   J. Biol. Chem. 0:0-0(2003).
DR   EMBL; AJ535114; CAD59380.1; -.
KW   Receptor; Signal.
FT   SIGNAL       1     18       Potential.
FT   CHAIN       19    487       fibroblast growth factor receptor-like
FT                               protein.
SQ   SEQUENCE   487 AA;  54099 MW;  FFD0132AD917FF94 CRC64;

  Query Match             5.8%;  Score 134;  DB 13;  Length 487;
  Best Local Similarity   21.2%;  Pred. No. 0.071;
  Matches   89;  Conservative   55;  Mismatches  149;  Indels  126;  Gaps   19;

Qy        117 EVQLLVFGLTANSDT---------HLLQGQSL----TLTLESP-PGSSPSVQCRSPRGKN 162
              :: ||: |: | ||:          ::  ||:    |: |  |  |  | :      |:
Db          4 QLALLLAGIVALSDSARGPPRIADKVIHRQSVRLGRTIKLLCPVEGDPPPLTMWMKDGRT 63

Qy        163 IQGGKT--------LSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKK 214
              |  | |        | : ::| :|:||: |             ::|:    :
Db         64 IHSGWTRFRILQQGLKIKEVESEDAGTYICKATNGFGSTNVNYTLIVIDDTSSGKNSQTP 123

Qy        215 EGEQVEFSFPLAFTVEKLTGSGELWWQ-------------AERASSS----------K 249
              ||   |:             ||: | |                |    ||          :
Db        124 EGSNGEYE----------DHSGKQWAQPRFTQPAKMRRRVIARPVGSSIRLKCVASGNPR 173

Qy        250 SWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLH 309
                ||:   || :    : ::      ||    | |    |:   ||  |  :  | |:::
Db        174 PDITWLKDNKPLMPHEIGEN------KKKKWTLNLKNLKPE--DSGKYTCRVFNKVGEIN 225

Qy        310 QEVNLVVMRATQLQK--------NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPV--W 359
                : |:: |: :          | | :  | ||             : ||   |||  |
Db        226 ATYKVEVIQRTRSKPILTGTHPVNTTVDYGGTTS----------FQCKVRSDVKPVIQW 274

Qy        360 VLNPEAG---MWQCLLSDSGQVLLESNIKVLPT---WSTP--------VHPRAS------ 399
              :   | |    :   :   ||  :     ||||   || |         :  ||
Db        275 LKRVEYGTESKYNSTIDVGGQKFV-----VLPTGEVWSRPDGSYLNKLMITRAKEEDAGM 329

Qy        400 --ALPAPPTG--------SALPDPQTASALPDPPAASALPAALAVISFLLGLGLGVACV 448
                | |   |         : ||||:  ||    | : |:||  :     ::|:  |   :
Db        330 YICLGANTMGYSFRSAFLTVLPDPKPPSAPVPPSSVSSLPWPV-----IIGIPAGAVFI 383


RESULT 63
Q63155
ID   Q63155      PRELIMINARY;      PRT;  1445 AA.
AC   Q63155;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Colorectal tumor suppressor.
GN   DCC.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97015074; PubMed=8861902;
RA   Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.Y.,
RA   Culotti J.G., Tessier-Lavigne M.;
RT   "Deleted in Colorectal Cancer (DCC) encodes a netrin receptor.";
RL   Cell 87:175-185(1996).
RN   [2]
RP   SEQUENCE OF 387-420 FROM N.A.
RX   MEDLINE=90100559; PubMed=2294591;
RA   Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA   Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA   Vogelstein B.;
RT   "Identification of a chromosome 18q gene that is altered in colorectal
RT   cancers.";
RL   Science 247:49-56(1990).
DR   EMBL; U68725; AAB41099.1; -.
DR   EMBL; M32291; AAA41086.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
```

```
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00408; IGc2; 3.
DR   PROSITE; PS50835; IG_LIKE; 4.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1445 AA;  157940 MW;  084F625954481988 CRC64;

  Query Match             5.8%;  Score 134;  DB 11;  Length 1445;
  Best Local Similarity   20.7%;  Pred. No. 0.32;
  Matches   75;  Conservative   48;  Mismatches  132;  Indels  108;  Gaps   14;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLND-RADSRRSLWDQGNF 92
              |||| | |          ||: : |                 ||    |||  :   |
Db        154 GDTVLLKCEVIGDPMPTIHWQKNQQ----------------DLNPIPGDSRVVVLPSG-- 195

Qy         93 PLIIKNLKIEDSDTYICEVED-------QKEEVQLL---------VFGLTANSDTHLLQGQ 137
               | |  |:  ||  | |   :        : ||::|        :: |   |:   ::|:
Db        196 ALQISRLQPGDSGVYRCSARNPASTRTGNEAEVRILSDPGLHRQLYFLQRPSNVIAIEGK 255

Qy        138 SLTL---TLESPPGS------SPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
                 |       || |          :| || :  :: ||  | :| :    ||||:|| |
Db        256 DAVLECCVSGYPPPSFTWLRGEEVIQLRSKK-YSLLGGSNLLISNVTDDDSGTYTCVVTY 314

Qy        189 NQKKVEFKIDIVVLA---FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERA 245
               : :    :: ||     |   | :|  |  :||                    | |
Db        315 KNENISASAELTVLVPPWFLNHPSNLYAYESMDIEF---------------------ECA 353

Qy        246 SSSKSWITFD-LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTL----- 299
               | |   | : :|| :|                       :|    |  |  || :
Db        354 VSGKPVPTVNWMKNGDV--------------------VIPSDYFQIVGGSNLRILGVVK 392

Qy        300 --------ALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKV 351
                     | : |       |:| :      ::      |::|: :: : : ::  ::
Db        393 SDEGFYQCVAENEAGNAQSSAQLIVPKPAIPSSSIL-----PSAPRDVVPVLVSSRFVRL 447

Qy        352 SKR 354
               | |
Db        448 SWR 450


RESULT 64
Q96AA2
ID   Q96AA2      PRELIMINARY;      PRT;  6620 AA.
AC   Q96AA2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Obscurin.
GN   OBSCN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=21342081; PubMed=11448995;
RA   Young P.W., Ehler E., Gautel M.;
RT   "Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor
RT   protein involved in sarcomere assembly.";
RL   J. Cell Biol. 154:123-136(2001).
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
DR   EMBL; AJ002535; CAC44768.1; -.
DR   Genew; HGNC:15719; OBSCN.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004812; F:tRNA ligase activity; IEA.
DR   GO; GO:0006418; P:amino acid activation; IEA.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000048; IQ_region.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000219; RhoGEF.
DR   InterPro; IPR001452; SH3.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 49.
DR   Pfam; PF00612; IQ; 1.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00621; RhoGEF; 1.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 17.
DR   SMART; SM00015; IQ; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR   PROSITE; PS50010; DH_2; 1.
DR   PROSITE; PS50835; IG_LIKE; 46.
DR   PROSITE; PS50096; IQ; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   6620 AA;  721665 MW;  C2AE8EB77B284452 CRC64;

  Query Match             5.7%;  Score 133.5;  DB 4;  Length 6620;
  Best Local Similarity   22.5%;  Pred. No. 2.9;
  Matches   91;  Conservative   59;  Mismatches  154;  Indels  101;  Gaps   20;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS-LWD---Q 89
              |: ||| |  | :    ||                        | ||   |:|  :|   :
Db       2841 GEDVELRCELS-RAGTPVHW---------------------LKDRKAIRKSQKYDVVCE 2877

Qy         90 GNFP-LIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLL-------QGQSL-T 140
              |    |:|:   ::|:  | ||||  |    | |    ||  |  |       :| :: |
Db       2878 GTMAMLVIRGASLKDAGEYTCEVEASKSTASLHV-EEKANCFTEELTNLQVEEKGTAVFT 2936

Qy        141 LTLESPPGSSPSVQCR--------SPRGKNIQGGKT--LSVSQLELQDSGTWTCTVLQNQ 190
                | |   : :|  |         | : :  | | |  |::| ||  || |:|| : | |
Db       2937 CKTEHP---AATVTWRKGLLELRASGKHQPSQEGLTLRLTISALEKADSDTYTCDIGQAQ 2993

Qy        191 KKVEFKI---------DIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEK---------- 231
               : :  :         |:  :  |: ||  ::       :  |: : ::|
Db       2994 SRAQLLVQGRRVHIIEDLEDVDVQEGSSATFRCRISPANYE-PVHWFLDKTPLHANELNE 3052

Qy        232 -----------------LTGSGELWWQA--ERASSSKSWITFDLKNKEVSVKRVTQDPKL 272
                              |  || ::::|  :|||::      :  |    |  |  :
Db       3053 IDAQPGGYHVLTLRQLALKDSGTIYFEAGDQRASAA-----LRVTEKPSVFSRELTDATI 3107

Qy        273 QMGKKLPLHL-TLPQALPQYAGSGNLTLALEAKTGKLHQ--EVNLVVMRAT-QLQKNLTC 328
               |: | |   |    :|       ||   |:    |:     |::  || |      |
Db       3108 TEGEDLTLVCETSTCDIPMCWTKDGKTLRGSARCQLSHEGHRAQLLITGATLQDSGRYKC 3167

Qy        329 EVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLS 373
               |  |  |  :   :::  :  :  :  | : ||   |   :|:||
Db       3168 EAGGACSSSI---VRVHARPVRFQEALKDLEVLEGGAATLRCVLS 3209


RESULT 65
Q8ISF3
ID   Q8ISF3      PRELIMINARY;      PRT;  2693 AA.
AC   Q8ISF3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   301KDa_2 protein.
GN   ISOF.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22269627; PubMed=12381307;
RA   Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
```

```
RA   Borodovsky M., Benian G.;
RT   "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT   Domains, Novel Regulation of Kinase Activity and Two New Possible
RT   Elastic Regions.";
RL   J. Mol. Biol. 323:533-549(2002).
DR   EMBL; AY130758; AAN61521.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 19.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 19.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 18.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
SQ   SEQUENCE   2693 AA;  299094 MW;  1CE4AD2011E8391A CRC64;

  Query Match             5.7%;  Score 133;  DB 5;  Length 2693;
  Best Local Similarity   20.7%;  Pred. No. 0.92;
  Matches   94;  Conservative   68;  Mismatches  161;  Indels  132;  Gaps   22;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              |||    |:|    || :     : | |                 | | :    ::: :  |:
Db       1100 VVLKTAGETATFTCQSYANPAAQVVW---------------LHNGKALQQTKSNYKTRLF 1144

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEV----QLLVFGLTANS-------------- 129
              |      |:|:|:  |    ||      :|  :|    || : |  |
Db       1145 DDNTATLVIENVTDELCGTYTAVANNQFGDVHTSAQLTISGSEAKKIAASLPYFIIELKP 1204

Qy        130 DTHLLQGQSLTLTLESPPGSSP----------SVQCRSPR------GKNIQGGKTLSVSQ 173
               ::::|   ||::::    ||           |    | |       | | |    | |
Db       1205 KINVVEG--ATLSIQADLNGSPIPEVVWLKDNSELVESDRIQMKCDGVNYQ----LLVRD 1258

Qy        174 LELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEF----------S 222
              : |:| ||:| |    : |:    :: |    |:  :  ||| ::||
Db       1259 VGLEDEGTYTITAENEKGKIRQNTEVSV---TKSKEVKEKKEKKKVEKKDEGKKKPGRPG 1315

Qy        223 FPL---AFTVEKLT--------GSGELWWQAERASSSKSWITF-DLKNKEVSVK------ 264
              |    |   |::|        |  : :    |    : |::    |: |: :|
Db       1316 LPRPSGASKTEQVTMAFDAPSEGPADSYEVERRCPDQREWVSCGSTKSLELEIKGLTPNT 1375

Qy        265 ----RVTQDPKLQMGKKLPLHLTLPQA----LPQYAGSGNLTLALEAKTGKLHQEVNLVV 316
                  ||    |  :|:   :  || |     ||:      |:: ::|         ::
Db       1376 EYIFRVAGKNKQGLGEWSEMTSTLKTASVGQAPQF------TISPQSK---------IIA 1420

Qy        317 MRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREK-------PVWVLN----PEA 365
              |  :  :  | |  :|    |:|   :  ::  ||         :||     |
Db       1421 NRDDEFE--IAVEFSGTPTP----SVKWYKENLQIVPDEKIDVATTSTSSILNLKSQEEN 1474

Qy        366 GMWQCLL-SDSGQVLLESNI----KVLPTWSTPVH 395
              | : ||: :: ||      :    |     ||| |
Db       1475 GTFNCLIENELGQASASCQVTIFNKPASLQSTPDH 1509


RESULT 66
Q8ISF4
ID   Q8ISF4      PRELIMINARY;      PRT;  2708 AA.
AC   Q8ISF4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   301KDa_1 protein.
GN   ISOF.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22269627; PubMed=12381307;
RA   Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA   Borodovsky M., Benian G.;
RT   "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT   Domains, Novel Regulation of Kinase Activity and Two New Possible
RT   Elastic Regions.";
RL   J. Mol. Biol. 323:533-549(2002).
DR   EMBL; AY130758; AAN61520.1; -.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 19.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 19.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 18.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
SQ   SEQUENCE   2708 AA;  300729 MW;  2BBEEBE856FA4571 CRC64;

  Query Match             5.7%;  Score 133;  DB 5;  Length 2708;
  Best Local Similarity   20.7%;  Pred. No. 0.92;
  Matches   94;  Conservative   68;  Mismatches  161;  Indels  132;  Gaps   22;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              |||    |:|    || :     : | |                 | | :    ::: :  |:
Db       1115 VVLKTAGETATFTCQSYANPAAQVVW---------------LHNGKALQQTKSNYKTRLF 1159

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEV----QLLVFGLTANS-------------- 129
              |      |:|:|:  |    ||      :|  :|    || : |  |
Db       1160 DDNTATLVIENVTDELCGTYTAVANNQFGDVHTSAQLTISGSEAKKIAASLPYFIIELKP 1219

Qy        130 DTHLLQGQSLTLTLESPPGSSP----------SVQCRSPR------GKNIQGGKTLSVSQ 173
               ::::|   ||::::    ||           |    | |       | | |    | |
Db       1220 KINVVEG--ATLSIQADLNGSPIPEVVWLKDNSELVESDRIQMKCDGVNYQ----LLVRD 1273

Qy        174 LELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEF----------S 222
              : |:| ||:| |    : |:    :: |    |:  :  ||| ::||
Db       1274 VGLEDEGTYTITAENEKGKIRQNTEVSV---TKSKEVKEKKEKKKVEKKDEGKKKPGRPG 1330

Qy        223 FPL---AFTVEKLT--------GSGELWWQAERASSSKSWITF-DLKNKEVSVK------ 264
              |    |   |::|        |  : :    |    : |::    |: |: :|
Db       1331 LPRPSGASKTEQVTMAFDAPSEGPADSYEVERRCPDQREWVSCGSTKSLELEIKGLTPNT 1390

Qy        265 ----RVTQDPKLQMGKKLPLHLTLPQA----LPQYAGSGNLTLALEAKTGKLHQEVNLVV 316
                  ||    |  :|:   :  || |     ||:      |:: ::|         ::
Db       1391 EYIFRVAGKNKQGLGEWSEMTSTLKTASVGQAPQF------TISPQSK---------IIA 1435

Qy        317 MRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREK-------PVWVLN----PEA 365
              |  : :  :  | |  :|    |:|   :  ::   ||          :||     |
```

```
Db       1436 NRDDEFE--IAVEFSGTPTP----SVKWYKENLQIVPDEKIDVATTSTSSILNLKSQEEN 1489

Qy        366 GMWQCLL-SDSGQVLLESNI----KVLPTWSTPVH 395
              | : ||: :: ||       :    |     ||| |
Db       1490 GTFNCLIENELGQASASCQVTIFNKPASLQSTPDH 1524


RESULT 67
Q8MNS0
ID   Q8MNS0      PRELIMINARY;      PRT;  2780 AA.
AC   Q8MNS0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   F12F3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Fulton B., Wohldmann P.;
RT   "The sequence of C. elegans cosmid F12F3.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80022; AAM29672.1; -.
DR   WormPep; F12F3.2a; CE30753.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 19.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 19.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 18.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Hypothetical protein; ATP-binding; Immunoglobulin domain; Repeat;
KW   Transferase.
SQ   SEQUENCE   2780 AA;  308646 MW;  046D057107B935C1 CRC64;

  Query Match             5.7%;  Score 133;  DB 5;  Length 2780;
  Best Local Similarity   20.7%;  Pred. No. 0.96;
  Matches   94;  Conservative   68;  Mismatches  161;  Indels  132;  Gaps   22;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              |||   |:|   || :    : |  |               |  | :    ::: :  |:
Db       1187 VVLKTAGETATFTCQSYANPAAQVVW--------------LHNGKALQQTKSNYKTRLF 1231

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEV----QLLVFGLTANS-------------- 129
              |     |:|:|:  |   ||     :|  :|    || : |  |
Db       1232 DDNTATLVIENVTDELCGTYTAVANNQFGDVHTSAQLTISGSEAKKIAASLPYFIIELKP 1291

Qy        130 DTHLLQGQSLTLTLESPPGSSP----------SVQCRSPR------GKNIQGGKTLSVSQ 173
               ::::|   ||::::    ||         |    | |      | | |    | |
Db       1292 KINVVEG--ATLSIQADLNGSPIPEVVWLKDNSELVESDRIQMKCDGVNYQ----LLVRD 1345

Qy        174 LELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEF-----------S 222
              : |:| ||:| |    : |:    :: |    |:  :  ||| ::||
Db       1346 VGLEDEGTYTITAENEKGKIRQNTEVSV---TKSKEVKEKKEKKKVEKKDEGKKKPGRPG 1402

Qy        223 FPL---AFTVEKLT--------GSGELWWQAERASSSKSWITF-DLKNKEVSVK------ 264
              |     |   |::|       |  : :    |    : |::    |: |: :|
Db       1403 LPRPSGASKTEQVTMAFDAPSEGPADSYEVERRCPDQREWVSCGSTKSLELEIKGLTPNT 1462

Qy        265 ----RVTQDPKLQMGKKLPLHLTLPQA----LPQYAGSGNLTLALEAKTGKLHQEVNLVV 316
                  ||     | :|:   :  ||  |      ||:      |:: ::|          ::
Db       1463 EYIFRVAGKNKQGLGEWSEMTSTLKTASVGQAPQF------TISPQSK---------IIA 1507

Qy        317 MRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREK-------PVWVLN----PEA 365
              |  : :  :  |  |  :|    |:|   :  ::   ||          :||     |
Db       1508 NRDDEFE--IAVEFSGTPTP----SVKWYKENLQIVPDEKIDVATTSTSSILNLKSQEEN 1561

Qy        366 GMWQCLL-SDSGQVLLESNI----KVLPTWSTPVH 395
              | : ||: :: ||       :    |     ||| |
Db       1562 GTFNCLIENELGQASASCQVTIFNKPASLQSTPDH 1596


RESULT 68
Q8MNS1
ID   Q8MNS1      PRELIMINARY;      PRT;  2808 AA.
AC   Q8MNS1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   F12F3.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Fulton B., Wohldmann P.;
RT   "The sequence of C. elegans cosmid F12F3.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80022; AAM29673.1; -.
DR   WormPep; F12F3.2b; CE30754.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
```

```
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 19.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 19.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 18.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Hypothetical protein; ATP-binding; Immunoglobulin domain; Repeat;
KW   Transferase.
SQ   SEQUENCE   2808 AA;  311979 MW;  02AFE0DAFE06FE12 CRC64;

  Query Match             5.7%;  Score 133;  DB 5;  Length 2808;
  Best Local Similarity   20.7%;  Pred. No. 0.97;
  Matches   94;  Conservative   68;  Mismatches  161;  Indels  132;  Gaps   22;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              |||   |:|   || :    : | |                 | | :    ::: :  |:
Db       1187 VVLKTAGETATFTCQSYANPAAQVVW--------------LHNGKALQQTKSNYKTRLF 1231

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEV----QLLVFGLTANS------------- 129
              |      |:|:|:  |   ||     :|  :|    || : |  |
Db       1232 DDNTATLVIENVTDELCGTYTAVANNQFGDVHTSAQLTISGSEAKKIAASLPYFIIELKP 1291

Qy        130 DTHLLQGQSLTLTLESPPGSSP----------SVQCRSPR------GKNIQGGKTLSVSQ 173
              ::::|   ||::::    ||          |    | |      | | |    | |
Db       1292 KINVVEG--ATLSIQADLNGSPIPEVVWLKDNSELVESDRIQMKCDGVNYQ----LLVRD 1345

Qy        174 LELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEF----------S 222
              : |:| ||:| |    : |:    :: |    |:  :  ||| ::||
Db       1346 VGLEDEGTYTITAENEKGKIRQNTEVSV---TKSKEVKEKKEKKKVEKKDEGKKKPGRPG 1402

Qy        223 FPL---AFTVEKLT--------GSGELWWQAERASSSKSWITF-DLKNKEVSVK------ 264
              |    |   |::|        |  : :    |    : |::    |: |: :|
Db       1403 LPRPSGASKTEQVTMAFDAPSEGPADSYEVERRCPDQREWVSCGSTKSLELEIKGLTPNT 1462

Qy        265 ----RVTQDPKLQMGKKLPLHLTLPQA----LPQYAGSGNLTLALEAKTGKLHQEVNLVV 316
                  ||    | :|:   :  ||  |     ||:      |:: ::|         ::
Db       1463 EYIFRVAGKNKQGLGEWSEMTSTLKTASVGQAPQF------TISPQSK---------IIA 1507

Qy        317 MRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREK-------PVWVLN----PEA 365
              |  : :  :  |  |  :|    |:|   :  ::   ||          :||     |
Db       1508 NRDDEFE--IAVEFSGTPTP----SVKWYKENLQIVPDEKIDVATTSTSSILNLKSQEEN 1561

Qy        366 GMWQCLL-SDSGQVLLESNI----KVLPTWSTPVH 395
              | : ||: :: ||      :    |     ||| |
Db       1562 GTFNCLIENELGQASASCQVTIFNKPASLQSTPDH 1596


RESULT 69
Q8ISF6
ID   Q8ISF6     PRELIMINARY;      PRT; 18519 AA.
AC   Q8ISF6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   2MDa_2 protein.
GN   ISOF.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22269627; PubMed=12381307;
RA   Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA   Borodovsky M., Benian G.;
RT   "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT   Domains, Novel Regulation of Kinase Activity and Two New Possible
RT   Elastic Regions.";
RL   J. Mol. Biol. 323:533-549(2002).
DR   EMBL; AY130758; AAN61518.1; -.
DR   PIR; E89066; E89066.
DR   PIR; T33247; T33247.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR001023; Hsp70.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 11.
DR   Pfam; PF00047; ig; 43.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000089; Hsp70; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00060; FN3; 11.
DR   SMART; SM00408; IGc2; 37.
DR   SMART; SM00406; IGv; 4.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS50835; IG_LIKE; 38.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   18519 AA;  2051869 MW;  6A8441C5D0BA7729 CRC64;

  Query Match             5.7%;  Score 133;  DB 5;  Length 18519;
  Best Local Similarity   20.7%;  Pred. No. 13;
  Matches   94;  Conservative   68;  Mismatches  161;  Indels  132;  Gaps   22;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
              |||   |:|   || :    : | |                 | | :    ::: :  |:
Db      16926 VVLKTAGETATFTCQSYANPAAQVVW--------------LHNGKALQQTKSNYKTRLF 16970

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEV----QLLVFGLTANS------------- 129
              |      |:|:|:  |   ||     :|  :|    || : |  |
Db      16971 DDNTATLVIENVTDELCGTYTAVANNQFGDVHTSAQLTISGSEAKKIAASLPYFIIELKP 17030

Qy        130 DTHLLQGQSLTLTLESPPGSSP----------SVQCRSPR------GKNIQGGKTLSVSQ 173
              ::::|   ||::::    ||          |    | |      | | |    | |
Db      17031 KINVVEG--ATLSIQADLNGSPIPEVVWLKDNSELVESDRIQMKCDGVNYQ----LLVRD 17084

Qy        174 LELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEF----------S 222
              : |:| ||:| |    : |:    :: |    |:  :  ||| ::||
Db      17085 VGLEDEGTYTITAENEKGKIRQNTEVSV---TKSKEVKEKKEKKKVEKKDEGKKKPGRPG 17141
```

```
Qy        223 FPL---AFTVEKLT--------GSGELWWQAERASSSKSWITF-DLKNKEVSVK------ 264
Db      17142 LPRPSGASKTEQVTMAFDAPSEGPADSYEVERRCPDQREWVSCGSTKSLELEIKGLTPNT 17201

Qy        265 ----RVTQDPKLQMGKKLPLHLTLPQA----LPQYAGSGNLTLALEAKTGKLHQEVNLVV 316
Db      17202 EYIFRVAGKNKQGLGEWSEMTSTLKTASVGQAPQF------TISPQSK---------IIA 17246

Qy        317 MRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREK-------PVWVLN----PEA 365
Db      17247 NRDDEFE--IAVEFSGTPTP----SVKWYKENLQIVPDEKIDVATTSTSSILNLKSQEEN 17300

Qy        366 GMWQCLL-SDSGQVLLESNI----KVLPTWSTPVH 395
Db      17301 GTFNCLIENELGQASASCQVTIFNKPASLQSTPDH 17335
```

```
RESULT 70
Q8ISF7
ID   Q8ISF7      PRELIMINARY;      PRT; 18534 AA.
AC   Q8ISF7;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   2MDa_1 protein.
GN   ISOF.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22269627; PubMed=12381307;
RA   Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA   Borodovsky M., Benian G.;
RT   "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT   Domains, Novel Regulation of Kinase Activity and Two New Possible
RT   Elastic Regions.";
RL   J. Mol. Biol. 323:533-549(2002).
DR   EMBL; AY130758; AAN61517.1; -.
DR   PIR; E89066; E89066.
DR   PIR; T33247; T33247.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR001023; Hsp70.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003596; Ig_v.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 11.
DR   Pfam; PF00047; ig; 43.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000089; Hsp70; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00060; FN3; 11.
DR   SMART; SM00408; IGc2; 37.
DR   SMART; SM00406; IGv; 4.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR   PROSITE; PS50835; IG_LIKE; 38.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   18534 AA;  2053504 MW;  BC8A682B943C8C0A CRC64;

  Query Match             5.7%;  Score 133;  DB 5;  Length 18534;
  Best Local Similarity   20.7%;  Pred. No. 14;
  Matches   94;  Conservative   68;  Mismatches  161;  Indels  132;  Gaps   22;

Qy         28 VVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLW 87
Db      16941 VVLKTAGETATFTCQSYANPAAQVVW--------------LHNGKALQQTKSNYKTRLF 16985

Qy         88 DQGNFPLIIKNLKIEDSDTYICEVEDQKEEV----QLLVFGLTANS-------------- 129
Db      16986 DDNTATLVIENVTDELCGTYTAVANNQFGDVHTSAQLTISGSEAKKIAASLPYFIIELKP 17045

Qy        130 DTHLLQGQSLTLTLESPPGSSP----------SVQCRSPR------GKNIQGGKTLSVSQ 173
Db      17046 KINVVEG--ATLSIQADLNGSPIPEVVWLKDNSELVESDRIQMKCDGVNYQ----LLVRD 17099

Qy        174 LELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEF----------S 222
Db      17100 VGLEDEGTYTITAENEKGKIRQNTEVSV---TKSKEVKEKKEKKKVEKKDEGKKKPGRPG 17156

Qy        223 FPL---AFTVEKLT--------GSGELWWQAERASSSKSWITF-DLKNKEVSVK------ 264
Db      17157 LPRPSGASKTEQVTMAFDAPSEGPADSYEVERRCPDQREWVSCGSTKSLELEIKGLTPNT 17216

Qy        265 ----RVTQDPKLQMGKKLPLHLTLPQA----LPQYAGSGNLTLALEAKTGKLHQEVNLVV 316
Db      17217 EYIFRVAGKNKQGLGEWSEMTSTLKTASVGQAPQF------TISPQSK---------IIA 17261

Qy        317 MRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREK-------PVWVLN----PEA 365
Db      17262 NRDDEFE--IAVEFSGTPTP----SVKWYKENLQIVPDEKIDVATTSTSSILNLKSQEEN 17315

Qy        366 GMWQCLL-SDSGQVLLESNI----KVLPTWSTPVH 395
Db      17316 GTFNCLIENELGQASASCQVTIFNKPASLQSTPDH 17350
```

```
RESULT 71
Q9W6V2
ID   Q9W6V2      PRELIMINARY;      PRT;   352 AA.
AC   Q9W6V2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurotractin-L.
GN   NTRA-L.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Bruemmendorf T.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=99264333; PubMed=10330412;
RA   Marg A., Sirim P., Spaltmann F., Plagge A., Kauselmann G., Buck F.,
RA   Rathjen F.G., Brummendorf T.;
RT   "Neurotractin, A novel neurite outgrowth-promoting Ig-like protein
RT   that interacts with CEPU-1 and LAMP.";
RL   J. Cell Biol. 145:865-876(1999).
```

```
DR   EMBL; AJ132999; CAB44446.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain.
SQ   SEQUENCE   352 AA;  37944 MW;  B0FA99F295FD8FA0 CRC64;

  Query Match             5.7%;  Score 132.5;  DB 13;  Length 352;
  Best Local Similarity   21.9%;  Pred. No. 0.059;
  Matches   76;  Conservative   55;  Mismatches  111;  Indels  105;  Gaps   17;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
Db         39 PGAAADSLVV--RKGDTAVLRCYLEDGAS-KGAWLNRSSIIFAG------------SDK 82

Qy         80 --ADSRRSL--WDQGNFPLIIKNLKIEDSDTYICEVEDQKE----------EVQLLVFGL 125
Db         83 WSVDPRVSIATANRREYSLQIQDVDVTDDGPYTCSVQTQHTPRTMQVHLTVQVSPKIFRI 142

Qy        126 TANSDTHLLQGQSLTLTLESPPGSSPSVQCR--SPRGKNIQGGKTLSVSQLELQDSGTWT 183
Db        143 --SSDIVVNEGSNVTLVCLATGKPEPSISWRHISPSAKPFESGQYLDIYGITRDQAGEYE 200

Qy        184 CTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGEL----- 238
Db        201 CS--------------------AENDVSVPDVKKVKVTVNFAPTIQELKSSGVMLGGNG 239

Qy        239 ---------------WWQAER-ASSSKSWITFDLKNKE----VSVKRVTQD--------- 269
Db        240 LIRCEGAGVPAPVFEWYRGERKLISGQQGIT--IKNYSTRSLLTVTNVTEEHFGNYTCVA 297

Qy        270 -PKLQM-GKKLPLHLTLPQALPQYAGSGN---------LTLALEAKT 305
Db        298 ANKLGMTNASLPLN---PPSTAQYGITGDAEVLFSCWYLVLTLSSLT 341


RESULT 72
Q9HCL6
ID   Q9HCL6      PRELIMINARY;      PRT;  1596 AA.
AC   Q9HCL6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein KIAA1556 (Fragment).
GN   KIAA1556.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20450683; PubMed=10997877;
RA   Nagase T., Kikuno R., Nakayama M., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.
RT   XVIII. The complete sequences of 100 new cDNA clones from brain which
RT   code for large proteins in vitro.";
RL   DNA Res. 7:273-281(2000).
DR   EMBL; AB046776; BAB13382.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 17.
DR   SMART; SM00408; IGc2; 7.
DR   PROSITE; PS50835; IG_LIKE; 15.
KW   Hypothetical protein; Immunoglobulin domain.
FT   NON_TER      1      1
FT   NON_TER    1596   1596
SQ   SEQUENCE   1596 AA;  175280 MW;  88937E0760C4E2EF CRC64;

  Query Match             5.7%;  Score 132.5;  DB 4;  Length 1596;
  Best Local Similarity   21.2%;  Pred. No. 0.48;
  Matches   90;  Conservative   59;  Mismatches  136;  Indels  139;  Gaps   20;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS-LWD---Q 89
Db        116 GEDVELRCELS-RAGTPVHW---------------------LKDRKAIRKSQKYDVVCE 152

Qy         90 GNFP-LIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLL-------QGQSL-T 140
Db        153 GTMAMLVIRGASLKDAGEYTCEVEASKSTASLHV-EEKANCFTEELTNLQVEEKGTAVFT 211

Qy        141 LTLESPPGSSPSVQCR--------SPRGKNIQGGKT--LSVSQLELQDSGTWTCTVLQNQ 190
Db        212 CKTEHP---AATVTWRKGLLELRASGKHQPSQEGLTLRLTISALEKADSDTYTCDIGQAQ 268

Qy        191 KKVEFKI---------DIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEK---------- 231
Db        269 SRAQLLVQGRRVHIIEDLEDVDVQEGSSATFRCRISPANYE-PVHWFLDKTPLHANELNE 327

Qy        232 -----------------LTGSGELWWQA--ERASSSKSWITFDLKNKEVSVKRVTQDPKL 272
Db        328 IDAQPGGYHVLTLRQLALKDSGTIYFEAGDQRASAA---------------LRVTEKPSV 372

Qy        273 QMGKKLPLHLTLPQALPQYAGSGNLTLALEAKT----------GKL-------------H 309
Db        373 FSRELTDATITEGE---------DLTLVCETSTCDIPVCWTKDGKTLRGSARCQLSHEGH 423

Qy        310 QEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQ 369
Db        424 RAQLLITGATLQDSGRYKCEAGGACSSSI---VRVHARPVRFQEALKDLEVLEGGAATLR 480

Qy        370 CLLS 373
Db        481 CVLS 484


RESULT 73
Q9VNP2
ID   Q9VNP2      PRELIMINARY;      PRT;  811 AA.
AC   Q9VNP2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG8779 protein.
GN   NRM OR CG8779.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
```

```
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003599; AAF51887.2; -.
DR   FlyBase; FBgn0005629; nrm.
DR   GO; GO:0005886; C:plasma membrane; IDA.
DR   InterPro; IPR002086; Aldehyde_dehydr.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00409; IG; 5.
DR   SMART; SM00408; IGc2; 4.
DR   PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR   PROSITE; PS50835; IG_LIKE; 7.
KW   Immunoglobulin domain.
SQ   SEQUENCE   811 AA;  90442 MW;  BAC689EA2C8E23F4 CRC64;

  Query Match             5.7%;  Score 132;  DB 5;  Length 811;
  Best Local Similarity   21.1%;  Pred. No. 0.2;
  Matches   83;  Conservative   68;  Mismatches  141;  Indels  102;  Gaps   20;

Qy         10 LLLVLQLALLPAAT-QGNKVVLGKKGDTVELTCTASQKKSIQFH----WKNSNQI--KIL 62
              |:||| |||: ::| | :  :  ::  :| | |    :|  : |    :|  ::|   :|
Db         26 LVLVLCLALVDSSTAQVDTTISQQESQSVVLPCPVDAEKCGKLHSLNWFKGDDRIAAMLL 85

Qy         63 GNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLV 122
              |:         : :|    | | :: :|  : |:||:||| | | |:|:
Db         86 GD------SNVTSVNKEFDERVTV-EQNPYRLVIKDLKIADEDIYLCDT----------T 128

Qy        123 FGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTW 182
              | :  :  :  | : | :  ||      |      :|  |: |    | :: : |
Db        129 FFIPEETCDN-FNGYRIELRVLVPP---TEVVILDAKGDRIKNGSV--VGPMQERQSLKA 182

Qy        183 TCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKL-TGSGELWWQ 241
              ||||   : : |                : : :  :::    |    |: | | : || |
Db        183 TCTVRNTRPQPE--------------VSWFRGTKRLTTYSPTHDLVDGLYTSTLELDWT 227

Qy        242 AERASSSKSWITFDLKNKEVSVKRVTQ---DPKLQM----------------GKKLPLHL 282
              |   ::     : : |  ::: ||       ||:                 | |: |
Db        228 LSREDLAQD---IECRVKSAAIQNVTVTKFSVDLQVRPTSIDINGVKHHTVQGSKVVLTC 284

Qy        283 TLPQALPQY------------AGSGNLT----LALEAKTGKLHQEVNLVVMRATQLQKN- 325
              :  | |              :|   :|     :||   |  | :  | :  ||: : :
Db        285 DIHGARPAVNLTWYNTTTIISSGENEITEVRSKSLEKSDGTFHTQSEL-IFNATRFENDR 343

Qy        326 -LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPV 358
               ||               ||   ::: ||||:
Db        344 VFRCEA-------------ENIVLQIN-REKPI 362


RESULT 74
Q24273
ID   Q24273      PRELIMINARY;      PRT;  1011 AA.
AC   Q24273;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NEUROMUSCULIN.
GN   NRM OR CG8779.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94000831; PubMed=8398154;
RA   Kania A., Han P.L., Kim Y.T., Bellen H.;
RT   "Neuromusculin, a Drosophila gene expressed in peripheral neuronal
RT   precursors and muscles, encodes a cell adhesion molecule.";
RL   Neuron 11:673-687(1993).
DR   EMBL; L23146; AAA03750.1; -.
DR   PIR; T13669; T13669.
DR   FlyBase; FBgn0005629; nrm.
DR   GO; GO:0005886; C:plasma membrane; IDA.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 8.
DR   SMART; SM00409; IG; 5.
DR   PROSITE; PS50835; IG_LIKE; 9.
KW   Cell adhesion.
SQ   SEQUENCE   1011 AA;  113125 MW;  AFD8A2A015D3AE63 CRC64;

  Query Match             5.7%;  Score 132;  DB 5;  Length 1011;
  Best Local Similarity   21.1%;  Pred. No. 0.28;
  Matches   83;  Conservative   68;  Mismatches  141;  Indels  102;  Gaps   20;

Qy         10 LLLVLQLALLPAAT-QGNKVVLGKKGDTVELTCTASQKKSIQFH----WKNSNQI--KIL 62
              |:||| |||: ::| | :  :  ::  :| | |    :|  : |    :|  ::|   :|
Db         26 LVLVLCLALVDSSTAQVDTTISQQESQSVVLPCPVDAEKCGKLHSLNWFKGDDRIAAMLL 85

Qy         63 GNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLV 122
              |:         : :|    | | :: :|  : |:||:||| | | |:|:
Db         86 GD------SNVTSVNKEFDERVTV-EQNPYRLVIKDLKIADEDIYLCDT----------T 128
```

```
Qy        123 FGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTW 182
Db        129 FFIPEETCDN-FNGYRIELRVLVPP---TEVVILDAKGDRIKNGSV--VGPMQERQSLKA 182

Qy        183 TCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKL-TGSGELWWQ 241
Db        183 TCTVRNTRPQPE--------------VSWFRGTKRLTTYSPTHDLVDGLYTSTLELDWT 227

Qy        242 AERASSSKSWITFDLKNKEVSVKRVTQ---DPKLQM---------------GKKLPLHL 282
Db        228 LSREDLAQD---IECRVKSAAIQNVTVTKFSVDLQVRPTSIDINGVKHHTVQGSKVVLTC 284

Qy        283 TLPQALPQY-----------AGSGNLT----LALEAKTGKLHQEVNLVVMRATQLQKN- 325
Db        285 DIHGARPAVNLTWYNTTTIISSGENEITEVRSKSLEKSDGTFHTQSEL-IFNATRFENDR 343

Qy        326 -LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPV 358
Db        344 VFRCEA-------------ENIVLQIN-REKPI 362


RESULT 75
Q9XT41
ID   Q9XT41      PRELIMINARY;      PRT;  1248 AA.
AC   Q9XT41;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neural cell adhesion molecule L1.
GN   L1CAM.
OS   Cercopithecus aethiops (Green monkey) (Grivet).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=9534;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=20259241; PubMed=10797421;
RA   Finckh U., Schroder J., Ressler B., Veske A., Gal A.;
RT   "Spectrum and detection rate of L1CAM mutations in isolated and
RT   familial cases with clinically suspected L1-disease.";
RL   Am. J. Med. Genet. 92:40-46(2000).
DR   EMBL; AF129167; AAD28610.1; -.
DR   HSSP; P20241; 1CFB.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00060; FN3; 4.
DR   SMART; SM00408; IGc2; 5.
DR   PROSITE; PS50835; IG_LIKE; 6.
KW   Immunoglobulin domain.
SQ   SEQUENCE   1248 AA;  138901 MW;  2438A93A29C3BD61 CRC64;

  Query Match             5.7%;  Score 132;  DB 6;  Length 1248;
  Best Local Similarity   19.4%;  Pred. No. 0.37;
  Matches   93;  Conservative   66;  Mismatches  191;  Indels  130;  Gaps   21;

Qy          8 RHLLLV----LQLALLPA--ATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKI 61
Db        404 RHGLLLANAYIYVVQLPAKILTADNQTYMAVQGSTAYLLCKAFGAPVPSVQWLDEDGTTV 463

Qy         62 LGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLL 121
Db        464 LQDERFF------------------PYANGTLGIRDLRANDTGRYLCLAANDQNNV-TI 503

Qy        122 VFGLTANSDTHLLQ---------GQSLTLTLES--PPGSSPSVQCRSPRGKNIQ------ 164
Db        504 VAHLKVKDATQITQGPRSAIEKKGSRVTFTCQASFDPSLQPSITWRGD-GRDLQELGDSD 562

Qy        165 -----GGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAF-----QKASSIVYKK 214
Db        563 KYFIEDGR-LVIHSLDYSDQGNYSCVASTELDVVESRAQLLVVGSPGPVPRLVLSDLHLL 621

Qy        215 EGEQVEFSFPLA-----------FTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
Db        622 TQSQVRVSWSPAEDHNAPIEKYDIEFEDKEMAPEKWYSLGKVPGNQTSTTLKLSPYVHYT 681

Qy        264 KRVTQDPKLQMGKKLPLHLTL--PQALPQ-----YAGSGNLTLALEAKTGKLHQEVNLVV 316
Db        682 FRVTAINKYGPGEPSPVSETVVTPEAAPEKNPVDVKGEGNET-------------TNMVI 728

Qy        317 MRATQLQKNLTCEVWGP------TSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQ- 369
Db        729 -------------TWKPLRWMDWNAPQVQYRVQ--------------WRPQGTRGPWQE 760

Qy        370 CLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALP 429
Db        761 QIVSD--PFLVVSNTSTF----VPYEIKVQAVNSQGKG---PEPQVTIGYSGEDYPQAIP 811


RESULT 76
Q7YQL7
ID   Q7YQL7      PRELIMINARY;      PRT;  1255 AA.
AC   Q7YQL7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   L1 cell adhesion molecule (Fragment).
GN   L1CAM.
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene Diversity Patterns at 10 X-Chromosomal Loci in Humans and
RT   Chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
DR   EMBL; AB102655; BAC81124.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   1255 AA;  139782 MW;  4FDAFCD6B12629C0 CRC64;

  Query Match             5.7%;  Score 132;  DB 6;  Length 1255;
  Best Local Similarity   19.4%;  Pred. No. 0.38;
  Matches   93;  Conservative   65;  Mismatches  192;  Indels  130;  Gaps   21;

Qy          8 RHLLLV----LQLALLPA--ATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKI 61
Db        407 RHGLLLANAYIYVVQLPAKILTADNQTYMAVQGSTAYLLCKAFGAPVPSVQWLDEDGTTV 466

Qy         62 LGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLL 121
Db        467 LQDERFF------------------PYANGTLGIRDLQANDTGRYFCLAANDQNNV-TI 506

Qy        122 VFGLTANSDTHLLQ---------GQSLTLTLES--PPGSSPSVQCRSPRGKNIQ------ 164
Db        507 VANLKVKDATQITQGPRSAIEKKGSRVTFTCQASFDPSLQPSITWRGD-GRDLQELGDSD 565

Qy        165 -----GGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAF-----QKASSIVYKK 214
Db        566 KYFIEDGR-LVIHSLDYSDQGNYSCVASTELDVVESRAQLLVVGSPGPVPRLVLSDLHLL 624

Qy        215 EGEQVEFSFPLA-----------FTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
Db        625 TQSQVRVSWSPAEDHNAPIEKYDIEFEDKEMAPEKWYSLGKVPGNQTSTTLKLSPYVHYT 684

Qy        264 KRVTQDPKLQMGKKLPLHLTL--PQALPQ-----YAGSGNLTLALEAKTGKLHQEVNLVV 316
Db        685 FRVTAINKYGPGEPSPVSETVVTPEAAPEKNPVDVKGEGNET-------------TNMVI 731
```

```
Qy        317 MRATQLQKNLTCEVWGP------TSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQ- 369
                                    | |       :|::   ::                |       | ||
Db        732 ------------TWKPLRWMDWNAPQVQYRVQ--------------WRPQGTRGPWQE 763

Qy        370 CLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALP 429
               ::||    |: ||        |   :  |: :   |   |:||                |:|
Db        764 QIVSD--PFLMVSNTSTF----VPYEIKVQAVNSQGKG---PEPQVTIGYSGEDYPQAIP 814


RESULT 77
Q8NHN3
ID   Q8NHN3      PRELIMINARY;      PRT;  2212 AA.
AC   Q8NHN3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Obscurin (Fragment).
GN   OBSCN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Young P., Ehler E., Gautel M.;
RT   "Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere
RT   assembly.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ314905; CAC85752.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 24.
DR   SMART; SM00409; IG; 25.
DR   SMART; SM00408; IGc2; 23.
DR   PROSITE; PS50835; IG_LIKE; 22.
KW   Immunoglobulin domain.
FT   NON_TER       1      1
FT   NON_TER    2212   2212
SQ   SEQUENCE   2212 AA;  241980 MW;  B6DFA263EB749465 CRC64;

  Query Match             5.7%;  Score 132;  DB 4;  Length 2212;
  Best Local Similarity   21.2%;  Pred. No. 0.83;
  Matches   90;  Conservative   59;  Mismatches  136;  Indels  140;  Gaps   20;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS-LWD---Q 89
              |: ||| |  | :     ||                                | ||   |:|  :|   :
Db        108 GEDVELRCELS-RAGTPVHW---------------------LKDRKAIRKSQKYDVVCE 144

Qy         90 GNFP-LIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLL-------QGQSL-T 140
              |    |:|:   ::|:  | ||||  |    | |    ||  |  |       :| :: |
Db        145 GTMAMLVIRGASLKDAGEYTCEVEASKSTASLHV-EEKANCFTEELTNLQVEEKGTAVFT 203

Qy        141 LTLESPPGSSPSVQCR--------SPRGKNIQGGKT--LSVSQLELQDSGTWTCTVLQNQ 190
                 | |    : :|  |        | : :  | | |  |::| ||  || |:|| : | |
Db        204 CKTEHP---AATVTWRKGLLELRASGKHQPSQEGLTLRLTISALEKADSDTYTCDIGQAQ 260

Qy        191 KKVEFKI----------DIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEK--------- 231
                : :  :          |:  :  |: ||  ::       :  |: : ::|
Db        261 SRAQLLVQAGRRVHIIEDLEDVDVQEGSSATFRCRISPANYE-PVHWFLDKTPLHANELN 319

Qy        232 -----------------LTGSGELWWQA--ERASSSKSWITFDLKNKEVSVKRVTQDPK 271
                              |  || ::::|  :|||::               |||: |
Db        320 EIDAQPGGYHVLTLRQLALKDSGTIYFEAGDQRASAA--------------LRVTEKPS 364

Qy        272 LQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKT----------GKL------------ 308
                :   :     :|  :          :|||  |  |           ||
Db        365 VFSRELTDATITEGE---------DLTLVCETSTCDIPVCWTKDGKTLRGSARCQLSHEG 415

Qy        309 HQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMW 368
               |:   |:     |       || |  |  :   ::: :  :  :  | : ||    |
Db        416 HRAQLLITGATLQDSGRYKCEAGGACSSSI---VRVHARPVRFQEALKDLEVLEGGAATL 472

Qy        369 QCLLS 373
                :|:||
Db        473 RCVLS 477


RESULT 78
Q23550
ID   Q23550      PRELIMINARY;      PRT;  6831 AA.
AC   Q23550; Q23020; Q27232;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   UNC-22 protein.
GN   UNC-22 OR ZK617.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   White S., Harris B.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 784-6831 FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=90044042; PubMed=2812002;
RA   Benian G.M., Kiff J.E., Neckelmann N., Moerman D.G., Waterson R.H.;
RT   "Sequence of an unusually large protein implicated in regulation of
RT   myosin activity in C. elegans.";
RL   Nature 342:45-50(1989).
RN   [3]
RP   SEQUENCE OF 784-6831 FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=93387664; PubMed=8397135;
RA   Benian G.M., L'Hernault S.W., Morris M.E.;
RT   "Additional sequence complexity in the muscle gene, unc-22, and its
RT   encoded protein, twitchin, of Caenorhabditis elegans.";
RL   Genetics 134:1097-1104(1993).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=a;
CC         IsoId=Q23550-1; Sequence=Displayed;
CC         Note=No experimental confirmation available;
CC       Name=b;
CC         IsoId=Q23551-1; Sequence=External;
CC         Note=No experimental confirmation available;
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; Z73897; CAA98064.1; -.
DR   EMBL; Z73899; CAA98064.1; JOINED.
DR   EMBL; Z73899; CAA98081.1; -.
DR   EMBL; Z73897; CAA98081.1; JOINED.
DR   EMBL; X15423; CAA33463.1; -.
DR   PIR; A88852; A88852.
DR   PIR; S57242; S57242.
DR   HSSP; Q63450; 1A06.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0004812; F:tRNA ligase activity; IEA.
DR   GO; GO:0006418; P:amino acid activation; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00041; fn3; 31.
DR   Pfam; PF00047; ig; 21.
DR   Pfam; PF00069; pkinase; 1.
```

```
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00060; FN3; 27.
DR   SMART; SM00408; IGc2; 6.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR   PROSITE; PS50835; IG_LIKE; 21.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Hypothetical protein; ATP-binding; Alternative splicing;
KW   Immunoglobulin domain; Kinase; Repeat;
KW   Serine/threonine-protein kinase; Transferase.
SQ   SEQUENCE   6831 AA;  752596 MW;  6CADE032963A52E1 CRC64;

  Query Match             5.7%;  Score 132;  DB 5;  Length 6831;
  Best Local Similarity   21.8%;  Pred. No. 4;
  Matches   97;  Conservative   66;  Mismatches  168;  Indels  114;  Gaps   23;

Qy          8 RHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
              ||  :|      |  :|  ||    : |   |    : | :    | :  :| | | :
Db        986 RHTFVV------PMKSQ--KV---NESDLATLETDVNDKDAEVVWWHDGKRIDIDGVK-- 1032

Qy         68 FLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFG--- 124
                    |:      ||        |||   :|||   | |  :| :   ||:|
Db       1033 ------FKVESSNRKRR---------LIINGARIEDHGEYKCTTKDDRTMAQLIVDAKNK 1077

Qy         125 -LTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ---GGK----------TLS 170
               : |  || :::   :||  ::    :| :  |:  || |    |||          ||
Db       1078 FIVALKDTEVIEKDDVTLMCQTKDTKTPGIWFRN--GKQISSMPGGKFETQSRNGTHTLK 1135

Qy         171 VSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYK------KEGEQVEFSFP 224
               : ::|: ::  :      :|  :    :: ||  :|   : :|      | ||      |
Db       1136 IGKIEMNEADVYEI----DQAGLRGSCNVTVLEAEKRPILNWKPKKIEAKAGEPCVVKVP 1191

Qy         225 LAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTL 284
                  | |  |:      :|   |:    | : ::: |: : :|   ::  |
Db       1192 FQI---KGTRRGD-----PKAQILKNGKPIDEEMRKL-VEVIIKDDVAEIVFK------- 1235

Qy         285 PQALPQYAGSGNLTLALEAKTGKLHQEVNLVVM------RATQLQKNLTCE----VWGPT 334
                 || | :|   | |    |       | |       :     ||:| |    |||
Db       1236 ---NPQLADTGKWALELGNSAGTALAPFELFVKDKPKPPKGPLETKNVTAEGLDLVWGTP 1292

Qy         335 SPKLMLSLK---LENKE------AKVSK-----------REKPVWVLNPEAGMWQCLLSD 374
                |     :|   :| :|      ||| :           :|   :    :| : :| |||
Db       1293 DPDEGAPVKAYIIEMQEGRSGNWAKVGETKGTDFKVKDLKEHGEYKFRVKA-LNECGLSD 1351

Qy         375 --SGQVLLESNIKVLPTWSTPVHPR 397
                 :|: :|  |      :  |  |:
Db       1352 PLTGESVLAKN-----PYGVPGKPK 1371


RESULT 79
Q23551
ID   Q23551      PRELIMINARY;      PRT;  7158 AA.
AC   Q23551;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein ZK617.1b (C. elegans UNC-22 protein)
DE   (Corresponding sequence ZK617.1b).
GN   ZK617.1B OR UNC-22.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   White S.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Harris B.R.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z73897; CAA98065.2; -.
DR   EMBL; Z73899; CAA98065.2; JOINED.
DR   EMBL; Z73897; CAA98082.2; JOINED.
DR   EMBL; Z73899; CAA98082.2; -.
DR   PIR; T27935; T27935.
DR   PDB; 1KOA; 30-APR-99.
DR   PDB; 1WIT; 23-DEC-96.
DR   PDB; 1WIU; 23-DEC-96.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0004812; F:tRNA ligase activity; IEA.
DR   GO; GO:0006418; P:amino acid activation; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 31.
DR   Pfam; PF00047; ig; 21.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00060; FN3; 32.
DR   SMART; SM00409; IG; 30.
DR   SMART; SM00408; IGc2; 24.
DR   SMART; SM00220; S_TKc; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
DR   PROSITE; PS50835; IG_LIKE; 21.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   7158 AA;  788947 MW;  561EF9DB13387505 CRC64;

  Query Match             5.7%;  Score 132;  DB 5;  Length 7158;
  Best Local Similarity   21.8%;  Pred. No. 4.3;
  Matches   97;  Conservative   66;  Mismatches  168;  Indels  114;  Gaps   23;

Qy          8 RHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
              ||  :|      |  :|  ||    : |   |    : | :    | :  :| | | :
Db       1313 RHTFVV------PMKSQ--KV---NESDLATLETDVNDKDAEVVWWHDGKRIDIDGVK-- 1359

Qy         68 FLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFG--- 124
                    |:      ||        |||   :|||   | |  :| :   ||:|
Db       1360 ------FKVESSNRKRR---------LIINGARIEDHGEYKCTTKDDRTMAQLIVDAKNK 1404

Qy         125 -LTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ---GGK----------TLS 170
               : |  || :::   :||  ::    :| :  |:  || |    |||          ||
Db       1405 FIVALKDTEVIEKDDVTLMCQTKDTKTPGIWFRN--GKQISSMPGGKFETQSRNGTHTLK 1462

Qy         171 VSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYK------KEGEQVEFSFP 224
               : ::|: ::  :      :|  :    :: ||  :|   : :|      | ||      |
Db       1463 IGKIEMNEADVYEI----DQAGLRGSCNVTVLEAEKRPILNWKPKKIEAKAGEPCVVKVP 1518
```

```
Qy        225 LAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTL 284
                  | |  |:      :|   |:    | : ::: |: : :|   ::  |
Db       1519 FQI---KGTRRGD-----PKAQILKNGKPIDEEMRKL-VEVIIKDDVAEIVFK------- 1562

Qy        285 PQALPQYAGSGNLTLALEAKTGKLHQEVNLVVM------RATQLQKNLTCE----VWGPT 334
                || | :|   | |    |       | |       :     ||:| |    |||
Db       1563 ---NPQLADTGKWALELGNSAGTALAPFELFVKDKPKPPKGPLETKNVTAEGLDLVWGTP 1619

Qy        335 SPKLMLSLK---LENKE------AKVSK----------REKPVWVLNPEAGMWQCLLSD 374
                |    :|   :| :|      ||| :          :|   :    :| : :| |||
Db       1620 DPDEGAPVKAYIIEMQEGRSGNWAKVGETKGTDFKVKDLKEHGEYKFRVKA-LNECGLSD 1678

Qy        375 --SGQVLLESNIKVLPTWSTPVHPR 397
                :|: :|  |      :  |  |:
Db       1679 PLTGESVLAKN-----PYGVPGKPK 1698


RESULT 80
Q25198
ID   Q25198      PRELIMINARY;      PRT;   848 AA.
AC   Q25198;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Receptor tyrosine kinase 90.
GN   HTK90.
OS   Hydra attenuata (Hydra) (Hydra vulgaris).
OC   Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
OC   Hydridae; Hydra.
OX   NCBI_TaxID=6087;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Miller M.A., Steele R.E.;
RT   "The Hydra Receptor Tyrosine Kinase, HTK90, is Expressed in a Subset
RT   of the Interstitial Cell Population.";
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U59448; AAB03389.1; -.
DR   HSSP; P08631; 1AD5.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   InterPro; IPR008266; Tyr_pkinase_AS.
DR   Pfam; PF00047; ig; 5.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 5.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW   ATP-binding; Immunoglobulin domain; Kinase; Transferase;
KW   Tyrosine-protein kinase.
SQ   SEQUENCE   848 AA;  95551 MW;  D0A52ED6A8760C07 CRC64;

  Query Match             5.7%;  Score 131.5;  DB 5;  Length 848;
  Best Local Similarity   20.0%;  Pred. No. 0.24;
  Matches   87;  Conservative   63;  Mismatches  181;  Indels  103;  Gaps   18;

Qy         37 VELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRAD-SRRSLWDQGNFPLI 95
              | : || |  :  |  |                  ||   :   :: ::  ::    | |
Db         48 VFVNCTTSPVEDAQIIW---------------NKGEENIPSYSEGTKDEVYVLPNNTLY 91

Qy         96 IKNLKIEDSDTYICEVEDQKEEV-QLLVF------GLTANSDTHLLQGQSLTLTLESPPG 148
              :|:|| ::: || |      |:: | :|       | :     :|||    :  | | |
Db         92 LKSLKKKEAGTYPCNAFLTLEKLFQTVVIEVAFLQKLNSIDPVKVLQGYVAEIVCEEPVG 151

Qy        149 S-SPSVQ------CRSPRGKNIQGGK-TLSVSQLELQDSGTWTCTVLQN--QKKVEFKID 198
                | |:         ||   ||     || : :  | |:| :||    :  ::  | |:
Db        152 KPQPKVKWLYNGATISPDNPNIDSSSWTLRIKKTMLSDAGNYTCVAFNSVAERTAEAKLT 211

Qy        199 IVVLAFQKASSIVYK-KEGEQVEFSF------PLAFTVEKLTGSGELWWQAERASSSKSW 251
              :||:     | |    ||||   |          ||            : |: :  : | :
Db        212 VVVVGDVDVSPISNSIKEGESATFKCESKSDPPL----------NVKWKRDIETCSTT- 259

Qy        252 ITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQE 311
                    :   || :||::  :   : :    ||           |       | |     :
Db        260 ------DNSQSVCKVTKEDVIPSDRVIISGGTLSIKNASVKDEGIYLCETELSTQISSAK 313

Qy        312 VNLVVMRATQLQKNL-----TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLN---- 362
              | | |    :: |:|     ||:     : |:     : : | :     |  | :||
Db        314 VQLNVFELMKVDKSLHNEIQTCDRLRQNNTKITCHFRGDGKYSIQWTRLSPNAILNSNMK 373

Qy        363 ------------PEAGMWQC--------------LLSDSGQVLLE-SNIKVL---PTW 390
                         : | ::|               :: :  | ::   ||       | |
Db        374 VVNDSIYISELKFEDMGQFKCEAVGEYNNATAYVNFIVYEHPQFIISPKNITAYIGEPAW 433

Qy        391 STPVHPRASALPAP 404
                || :    | |
Db        434 ---VHCQGKGFPKP 444


RESULT 81
Q7SX76
ID   Q7SX76      PRELIMINARY;      PRT;   483 AA.
AC   Q7SX76;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FGF receptor-like protein precursor (Hypothetical protein).
GN   FGFRL1.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Trueb B., Zhuang L., Taeschler S., Wiedemann M.;
RT   "Characterization of FGFRL1, a novel FGF receptor preferentially
RT   expressed in skeletal tissues.";
RL   J. Biol. Chem. 278:0-0(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
```

```
RA   Strausberg R.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ574916; CAE00573.1; -.
DR   EMBL; BC053245; AAH53245.1; -.
KW   Hypothetical protein; Receptor; Signal.
FT   SIGNAL        1     19       Potential.
FT   CHAIN        20    483       FGF receptor-like protein.
SQ   SEQUENCE   483 AA;  53937 MW;  3C2B34E5D57C332C CRC64;

  Query Match             5.6%;  Score 131;  DB 13;  Length 483;
  Best Local Similarity   21.6%;  Pred. No. 0.12;
  Matches   85;  Conservative   48;  Mismatches  128;  Indels  132;  Gaps   20;

Qy        140 TLTLESP-PGSSPSVQCRSPRGKNIQGG--------KTLSVSQLELQDSGTWTCTVLQNQ 190
              |: |: |  |  | :   :  |:||  |        : | : ::|  |:||: |
Db         41 TMKLQCPVEGDPPPLIMWTKDGRNIHSGWMRFRVLQQALRIKEVEADDAGTFICKATNGF 100

Qy        191 KKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSF------------------------PLA 226
              |     ::|:   |    :  ||  |:|                          |:
Db        101 GSVNINYTLIVIDDSSAGREGARPAGE-TEYSTDLTGKLVRPRFTQPAKMRKRVIARPVG 159

Qy        227 FTVE-KLTGSGE-----LWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPL 280
              :|  | | ||      :| :  |           |  :||             |:|
Db        160 SSVRLKCTASGNPRPDIVWLKDSR----------PLTPEEVG----------EGRKKKW 198

Qy        281 HLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQK--------NLTCEVWG 332
              |:|    |::  ||  |  :  : |:::    : |:: |  :          | | :  |
Db        199 TLSLKNLTPEH--SGKYTCHVSNRAGEINATYKVEVIQRTNSKPILTGTHPVNTTVDYGG 256

Qy        333 PTSPKLMLSLKLENKEAKVSKREKPV--WVLNPEAGMWQCLLSDSGQVLLESNIK----- 385
              ||              : ||   |||  |:   | |          |:   | |:
Db        257 TTS-----------FQCKVRSDVKPVIQWLKRVEPG---------GEGKYNSTIEVGDHH 296

Qy        386 --VLPT---WSTP--------VHPRAS--------ALPAPPTG--------SALPDPQTA 416
              ||||   || |          :  ||          | |   |         : ||||:
Db        297 FVVLPTGDVWSRPDGSYLNKLLITRAKEEDAGMYICLGANTMGYSFRSAFLTVLPDPKP- 355

Qy        417 SALPDPPAASALPAALAVISFLLGLGLGVACVL 449
                |  |  | || :|     ::|:  |:  :|
Db        356 ---PFSPIPSVLPPSLP-WPVIIGIPAGIVFIL 384


RESULT 82
Q80TU8
ID   Q80TU8      PRELIMINARY;      PRT;   827 AA.
AC   Q80TU8;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MKIAA0657 protein (Fragment).
GN   MKIAA0657.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22579291; PubMed=12693553;
RA   Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA   Nakajima D., Nagase T., Ohara O., Koga H.;
RT   "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT   II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT   cDNAs identified by screening of terminal sequences of cDNA clones
RT   randomly sampled from size-fractionated libraries.";
RL   DNA Res. 10:35-48(2003).
DR   EMBL; AK122340; BAC65622.2; -.
FT   NON_TER       1      1
SQ   SEQUENCE   827 AA;  90465 MW;  ADD3369FB3A21DE6 CRC64;

  Query Match             5.6%;  Score 131;  DB 11;  Length 827;
  Best Local Similarity   20.5%;  Pred. No. 0.25;
  Matches  102;  Conservative   58;  Mismatches  166;  Indels  172;  Gaps   22;

Qy         18 LLPAATQGNKVVLGKKGDTVELTCTASQKKSI-------QFHW--KNSNQIKILG----- 63
              :||||         :  |  |  | |  :::         | |    |::: :
Db         19 VLPAA---------RPSDGGEFQCVAGDERAYFTVTITDVFSWIVYPSSEVHVAAVRLER 69

Qy         64 ----------------NQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDT 106
                            |  : : |: | :: |: |          |:: ::::|||
Db         70 VVLTCELCRPWAEVRWTKDGEEVVESPALLLEKEDTIRR--------LVLPSVQLEDSGE 121

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
              |:||: |                    :  | |:|:  ||     |:   |     |
Db        122 YLCEIHD--------------------ESASFTITVTEPP-----VRIIYP-----QDE 150

Qy        167 KTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLA 226
              ||    ||       || : :    | :  |   |  ::: ::| : :| :     | |
Db        151 VTLHAVSLECV---VLTCELSREDAPVRWYKD--GLEVEESEALVLQSDGPRRRLVLPAA 205

Qy        227 ------------------FTV------EKLT--------------GSGELWWQAERASSS 248
                               |||      |::                |  ||  :   | |
Db        206 QPEDGGEFVCDAGDDSAFFTVTVTAPPERIVHPAARSLDLQFGAPGHVELRCEVAPAGSQ 265

Qy        249 KSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLH-LTLPQALPQYAGSGNLTLALEAKTGK 307
              |    |   || |        ||:| : |   |||| | |: ||        || |
Db        266 VRWYKDGL---EVEV-----SDALQLGAEGPARTLTLPHAQPEDAGEYVCETRDEAVTFN 317

Qy        308 LHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKV---------------- 351
              :      |   | :   |  | | |  |  | ::|| :|    |:|
Db        318 VSLAELPVQFLAPEAAPNPLCVV--PGEP-VVLSCELSRASAQVFWSHNGSPVQQGEGLE 374

Qy        352 SKREKPVWVLNPEA------GMWQCLLSDS-GQVLLESNIKVLPTWSTPVHPRASALPAP 404
              : | |  :|  :|      |:: |    | |   |  |::|                |
Db        375 LRAEGPRRILCIQAADLAHTGVYTCQSGASPGAPSLSFNVQV--------------AEP 419

Qy        405 PTGSALPDPQTASALPDP 422
              |     |:    |    |
Db        420 PVRVVAPEAAQTSVRSTP 437


RESULT 83
Q7Z3Z9
ID   Q7Z3Z9      PRELIMINARY;      PRT;  1255 AA.
AC   Q7Z3Z9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   L1 cell adhesion molecule (Fragment).
GN   L1CAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene Diversity Patterns at 10 X-Chromosomal Loci in Humans and
RT   Chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
DR   EMBL; AB102653; BAC81122.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   1255 AA;  139772 MW;  D6068987CD688B27 CRC64;

  Query Match             5.6%;  Score 131;  DB 4;  Length 1255;
  Best Local Similarity   19.2%;  Pred. No. 0.45;
  Matches   92;  Conservative   66;  Mismatches  192;  Indels  130;  Gaps   21;

Qy          8 RHLLLV----LQLALLPA--ATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKI 61
              || ||:    : :  |||   |  |:  :  :| |  | | |         | : :   :
Db        407 RHGLLLANAYIYVVQLPAKILTADNQTYMAVQGSTAYLLCKAFGAPVPSVQWLDEDGTTV 466

Qy         62 LGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLL 121
              | ::  |                    |  | |::|:  |:  | |   : :  | ::
Db        467 LQDERFF------------------PYANGTLGIRDLQANDTGRYFCLAANDQNNVTIM 507
```

```
Qy        122 VFGLTANSDTHLLQ---------GQSLTLTLES--PPGSSPSVQCRSPRGKNIQ------ 164
Db        508 A-NLKVKDATQITQGPRSTIEKKGSRVTFTCQASFDPSLQPSITWRGD-GRDLQELGDSD 565

Qy        165 -----GGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAF-----QKASSIVYKK 214
Db        566 KYFIEDGR-LVIHSLDYSDQGNYSCVASTELDVVESRAQLLVVGSPGPVPRLVLSDLHLL 624

Qy        215 EGEQVEFSFPLA-----------FTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
Db        625 TQSQVRVSWSPAEDHNAPIEKYDIEFEDKEMAPEKWYSLGKVPGNQTSTTLKLSPYVHYT 684

Qy        264 KRVTQDPKLQMGKKLPLHLTL--PQALPQ-----YAGSGNLTLALEAKTGKLHQEVNLVV 316
Db        685 FRVTAINKYGPGEPSPVSETVVTPEAAPEKNPVDVKGEGNET------------TNMVI 731

Qy        317 MRATQLQKNLTCEVWGP------TSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQ- 369
Db        732 ------------TWKPLRWMDWNAPQVQYRVQ--------------WRPQGTRGPWQE 763

Qy        370 CLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALP 429
Db        764 QIVSD--PFLVVSNTSTF----VPYEIKVQAVNSQGKG---PEPQVTIGYSGEDYPQAIP 814


RESULT 84
Q7YQL8
ID   Q7YQL8      PRELIMINARY;      PRT;  1255 AA.
AC   Q7YQL8;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   L1 cell adhesion molecule (Fragment).
GN   L1CAM.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22763540; PubMed=12777533;
RA   Kitano T., Schwarz C., Nickel B., Paabo S.;
RT   "Gene Diversity Patterns at 10 X-Chromosomal Loci in Humans and
RT   Chimpanzees.";
RL   Mol. Biol. Evol. 20:1281-1289(2003).
DR   EMBL; AB102654; BAC81123.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   1255 AA;  139742 MW;  235D47715A3DD6D9 CRC64;

  Query Match             5.6%;  Score 131;  DB 6;  Length 1255;
  Best Local Similarity   19.4%;  Pred. No. 0.45;
  Matches   93;  Conservative   65;  Mismatches  192;  Indels  130;  Gaps   21;

Qy          8 RHLLLV----LQLALLPA--ATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKI 61
Db        407 RHGLLLANAYIYVVQLPAKILTADNQTYMAVQGSTAYLLCKAFGAPVPSVQWLDEDGTTV 466

Qy         62 LGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLL 121
Db        467 LQDERFF------------------PYANGTLGIRDLQANDTGRYFCLAANDQNNV-TI 506

Qy        122 VFGLTANSDTHLLQ---------GQSLTLTLES--PPGSSPSVQCRSPRGKNIQ------ 164
Db        507 VANLKVKDATQITQGPRSAIEKKGSRVTFTCQASFDPSLQPSITWRGD-GRDLQELGDSD 565

Qy        165 -----GGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAF-----QKASSIVYKK 214
Db        566 KYFIEDGR-LVIHSLDYSDQGNYSCVASTELDVVESRAQLLVVGSPGPVPRLVLSDLHLL 624

Qy        215 EGEQVEFSFPLA-----------FTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
Db        625 TQSQVRVSWSPAEDHNAPIEKYDIEFEDKEMAPEKWYSLGKVPGNQTSTTLKLSPYVHYT 684

Qy        264 KRVTQDPKLQMGKKLPLHLTL--PQALPQ-----YAGSGNLTLALEAKTGKLHQEVNLVV 316
Db        685 FRVTAINKYGPGEPSPVSETVVTPEAAPEKNPMDVKGEGNET------------TNMVI 731

Qy        317 MRATQLQKNLTCEVWGP------TSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQ- 369
Db        732 ------------TWKPLRWMDWNAPQVQYRVQ--------------WRPQGTRGPWQE 763

Qy        370 CLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALP 429
Db        764 QIVSD--PFLVVSNTSTF----VPYEIKVQAVNSQGKG---PEPQVTIGYSGEDYPQAIP 814


RESULT 85
Q8NDA2
ID   Q8NDA2      PRELIMINARY;      PRT;  1340 AA.
AC   Q8NDA2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP434P0216.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Koehrer K., Beyer A., Mewes H.W., Weil B., Wiemann S.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL834139; CAD38854.1; -.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00047; ig; 15.
DR   SMART; SM00408; IGc2; 14.
DR   PROSITE; PS50835; IG_LIKE; 14.
KW   Hypothetical protein; Immunoglobulin domain.
FT   NON_TER       1       1
SQ   SEQUENCE   1340 AA;  143889 MW;  7AC26BE866DB7FBF CRC64;

  Query Match             5.6%;  Score 131;  DB 4;  Length 1340;
  Best Local Similarity   19.7%;  Pred. No. 0.49;
  Matches   80;  Conservative   58;  Mismatches  163;  Indels  106;  Gaps   16;

Qy         12 LVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTK 71
Db        929 LVVQVP--PVIENGLPDLSTTEGSHAFLPCKARGSPEPNITWDKDGQ-PVSGAEGKFTI- 984

Qy         72 GPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED----QKEEVQLLVFGL-- 125
Db        985 ----------------QPSGELLVKNLEGQDAGTYTCTAENAVGRARRRVHLTILVLPV 1027

Qy        126 --TANSDTHLLQGQSLTLTLESPPGSSPSVQCR---SPRGKNI---QGGKTLSVSQLELQ 177
Db       1028 FTTLPGDRSLRLGDRLWLRCAARGSPTPRIGWTVNDRPVTEGVSEQDGGSTLQRAAVSRE 1087

Qy        178 DSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGS-- 235
Db       1088 DSGTYVCWA-ENR-----------VGRTQAVSFVHVKEAPVLQ-GEAFSYLVEPVGGSIQ 1134

Qy        236 ----------GELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLP 285
Db       1135 LDCVVRGDPVPDIHWIKDGLPLRGSHLRHQLQNGSLTIHRTERD---------------- 1178

Qy        286 QALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKN------------LTCEVWG 332
Db       1179 -------DAGRYQCLAENEMGVAKKVVILVLQSAPVFQVEPQDMTVRSGDDVALRCQATG 1231

Qy        333 PTSPKLMLSLKLENKEAKVSKR-----EKPVWVLNPE---AGMWQCL 371
Db       1232 EPTP--TIEWLQAGQPLRASRRLRTLPDGSLWLENVETGDAGTYDCV 1276
```

```
RESULT 86
Q8I0L3
ID   Q8I0L3      PRELIMINARY;      PRT;  5175 AA.
AC   Q8I0L3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   C. elegans him-4 protein (corresponding sequence F15G9.4a).
GN   F15G9.4 OR HIM-4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sulston J.E.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Kershaw J.K.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z47068; CAA87335.1; -.
DR   EMBL; Z47070; CAA87335.1; JOINED.
DR   EMBL; Z47070; CAA87344.1; -.
DR   EMBL; Z47068; CAA87344.1; JOINED.
DR   PIR; T20992; T20992.
DR   WormPep; F15G9.4a; CE18595.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000515; BPD_transp.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 47.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00409; IG; 45.
DR   SMART; SM00408; IGc2; 47.
DR   SMART; SM00406; IGv; 12.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50835; IG_LIKE; 47.
SQ   SEQUENCE   5175 AA;  568471 MW;  4B2561803BBC62A4 CRC64;

  Query Match             5.6%;  Score 131;  DB 5;  Length 5175;
  Best Local Similarity   19.5%;  Pred. No. 3.2;
  Matches   85;  Conservative   68;  Mismatches  195;  Indels   88;  Gaps   18;

Qy         33 KGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNF 92
              : ||: : |  : :      |              |    | | |    | |:   |:
Db       3495 ENDTITMDCGVTSRPLPSISW-------------FRGDKPVYLYD----RYSISPDGSH 3536

Qy         93 PLIIKNLKIEDSDTYICEVEDQ--KEEVQLLVFGLT-ANSDTHLLQGQSL-----TLTLE 144
              : |   |: |   |||   ::    :: |::  |      |   : | |     |: ||
Db       3537 -ITINKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIGNPLAIVARTIYLE 3595

Qy        145 SPPGSSP----------SVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
              |   |            :    |   | :|  :  ::: | | :|||     |
Db       3596 CPISGIPQPDVIWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGGKA 3655

Qy        194 EFKIDIVVLAFQK----ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSK 249
              : ||:  :    :    |:||: :  : |:    :      :: | :  :
Db       3656 SHDFSLDVLSPPEFDIHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRA--- 3712

Qy        250 SWITFDL-KNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKL 308
              :  ||  | ::|           |:|    ||: ||  : ||          |    |
Db       3713 --LDGDLTDNVDIS----------DDGRK----LTISQASLENAGLYTCIALNRAGEASL 3756

Qy        309 HQEVNLVVMRATQLQKN------------LTCEVWGPTSPKLMLSLK-----LENKEAK 350
              :| ::      : :|             : | | |   | :   ||      :::  :
Db       3757 EFKVEILSPPVIDISRNDVQPQVAVNQPTIMRCAVTGHPFPSIKW-LKNGKEVTDDENIR 3815

Qy        351 VSKREKPVWVLNPE---AGMWQCLL-SDSGQVLLESNIKVLPTWSTPVHPRASALPAPPT 406
              : :: : : :|  :   || | |:  :|:|   ||    ||  :: ||    |  |
Db       3816 IVEQGQVLQILRTDSDHAGKWSCVAENDAGVKELE---MVLDVFTPPVVSVKSDNPIKAL 3872

Qy        407 GSALPDPQTASALPDP 422
              |  :     ||  | |
Db       3873 GETITLFCNASGNPYP 3888


RESULT 87
O76518
ID   O76518      PRELIMINARY;      PRT;  5198 AA.
AC   O76518; Q10036;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hemicentin precursor (C. elegans him-4 protein) (corresponding
DE   sequence F15G9.4b).
GN   F15G9.4 OR HIM-4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   PubMed=11222143;
RA   Vogel B.E., Hedgecock E.M.;
RT   "Hemicentin, a conserved extracellular member of the immunoglobulin
RT   superfamily, organizes epithelial and other cell attachments into
RT   oriented line-shaped junctions.";
RL   Development 128:883-894(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Sulston J.E.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Kershaw J.K.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF074901; AAC26792.1; -.
DR   EMBL; Z47068; CAA87336.1; -.
DR   EMBL; Z47070; CAA87336.1; JOINED.
DR   EMBL; Z47070; CAA87345.1; -.
DR   EMBL; Z47068; CAA87345.1; JOINED.
DR   PIR; T43290; T43290.
DR   HSSP; P00736; 1APQ.
DR   WormPep; F15G9.4b; CE18596.
DR   GO; GO:0016020; C:membrane; IEA.
```

```
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005215; F:transporter activity; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000515; BPD_transp.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00047; ig; 47.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00408; IGc2; 44.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50835; IG_LIKE; 47.
KW   EGF-like domain; Immunoglobulin domain; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   5198       HEMICENTIN.
SQ   SEQUENCE   5198 AA;  570809 MW;  DA8511FF2B58D37B CRC64;

  Query Match             5.6%;  Score 131;  DB 5;  Length 5198;
  Best Local Similarity   19.5%;  Pred. No. 3.3;
  Matches   85;  Conservative   68;  Mismatches  195;  Indels   88;  Gaps   18;

Qy         33 KGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNF 92
              : ||: : |  : :      |                |    | | |    | |:  |:
Db       3495 ENDTITMDCGVTSRPLPSISW-------------FRGDKPVYLYD----RYSISPDGSH 3536

Qy         93 PLIIKNLKIEDSDTYICEVEDQ--KEEVQLLVFGLT-ANSDTHLLQGQSL-----TLTLE 144
               : |   |: |   |||   ::   :: |::  |     |   : |  |     |: ||
Db       3537 -ITINKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIGNPLAIVARTIYLE 3595

Qy        145 SPPGSSP-----------SVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
              |    |           :     |    | :|  :  ::: | | :|||      |
Db       3596 CPISGIPQPDVIWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGGKA 3655

Qy        194 EFKIDIVVLAFQK----ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSK 249
                : ||:  :     :    |:||: :  : |:     :       :: |  :  :
Db       3656 SHDFSLDVLSPPEFDIHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRA--- 3712

Qy        250 SWITFDL-KNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKL 308
               :  ||  | ::|             |:|    ||: ||  : ||          |    |
Db       3713 --LDGDLTDNVDIS----------DDGRK----LTISQASLENAGLYTCIALNRAGEASL 3756

Qy        309 HQEVNLVVMRATQLQKN------------LTCEVWGPTSPKLMLSLK-----LENKEAK 350
               :| ::       : :|             : | | |   | :   ||      :::  :
Db       3757 EFKVEILSPPVIDISRNDVQPQVAVNQPTIMRCAVTGHPFPSIKW-LKNGKEVTDDENIR 3815

Qy        351 VSKREKPVWVLNPE---AGMWQCLL-SDSGQVLLESNIKVLPTWSTPVHPRASALPAPPT 406
              : :: : : :|  :   || | |:  :|:|   ||    ||  :: ||    |   |
Db       3816 IVEQGQVLQILRTDSDHAGKWSCVAENDAGVKELE---MVLDVFTPPVVSVKSDNPIKAL 3872

Qy        407 GSALPDPQTASALPDP 422
              |  :     ||  | |
Db       3873 GETITLFCNASGNPYP 3888


RESULT 88
Q98918
ID   Q98918      PRELIMINARY;      PRT;  4162 AA.
AC   Q98918; P87479;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Connectin/titin (Fragment).
GN   TITIN.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Breast muscle;
RX   MEDLINE=96254045; PubMed=8660363;
RA   Yajima H., Ohtsuka H., Kawamura Y., Kume H., Murayama T., Abe H.,
RA   Kimura S., Maruyama K.;
RT   "A 11.5-kb 5'-terminal cDNA sequence of chicken breast muscle
RT   connectin/titin reveals its Z line binding region.";
RL   Biochem. Biophys. Res. Commun. 223:160-164(1996).
RN   [2]
RP   SEQUENCE OF 1767-1871 FROM N.A.
RX   MEDLINE=96365546; PubMed=8769723;
RA   Turnacioglu K.T., Mittal B., Sanger J.M., Sanger J.W.;
RT   "Partial characterization of zeugmatin indicates that it is part of
RT   the Z-band region of titin.";
RL   Cell Motil. Cytoskeleton 34:108-121(1996).
CC   -!- SIMILARITY: CONTAINS 33 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
DR   EMBL; D83390; BAA11908.1; -.
DR   EMBL; U64829; AAC60019.1; -.
DR   PIR; T42633; T42633.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 33.
DR   SMART; SM00408; IGc2; 14.
DR   PROSITE; PS50835; IG_LIKE; 29.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Immunoglobulin domain.
FT   NON_TER    4162   4162
SQ   SEQUENCE   4162 AA;  464971 MW;  DB0905C907686649 CRC64;

  Query Match             5.6%;  Score 130;  DB 13;  Length 4162;
  Best Local Similarity   22.4%;  Pred. No. 2.9;
  Matches  103;  Conservative   55;  Mismatches  159;  Indels  142;  Gaps   24;

Qy         11 LLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHW-KNSNQI------KILG 63
              | ||: |::       ||     ||:  | ||       :    | |: |::      ||
Db       3551 LTVLEPAVIVEKPGPVKVT---AGDSCTLECTVDGTPELTARWFKDGNELSTDHKYKI-- 3605

Qy         64 NQGSFLTK--GPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLL 121
                 ||  |  |   ||        | | | :   :||   : | |   :| |:
Db       3606 ---SFFNKVSGLKILN------AGLEDSGEYTFEVKNSVGKSSCTASLQVSDR------- 3649

Qy        122 VFGLTANSDTHLLQGQSLTLTLE---SPPGSSPSVQCR---SP--------RGKNIQGGK 167
                       :   | |  |:      |||  ::|:   ||          |: | |
Db       3650 ------------IMPPSFTRKLKETYGQLGSSAVLECKVYGSPPILVSWFHDGQEITSGD 3697

Qy        168 -----------TLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKE- 215
                        :| |: |:  | ||::||        |       |: ::  | | | |
Db       3698 KYQATLTDNTCSLKVNGLQESDMGTYSCTATNVAGSDECS---AFLSVREPPSFVKKPEP 3754

Qy        216 -----GEQVEFSFPLAFTVEKLTGSGELWW---QAERASSSKSWITFDLKNKEVSVKRVT 267
                  || : |:     :: | :   |: |      | |   |  ||
Db       3755 FNVLSGENITFT-----SIVKGSPPLEVKWFRGSIELAPGHKCNITL------------ 3796

Qy        268 QD--PKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNL-------VVMR 318
               ||   :|::    ||             ||: |  :  : ||:    :|        ||:
Db       3797 QDSVAELELFDVQPLQ-----------SGDYTCQVSNEAGKISCTTHLFVKEPAKFVMK 3844

Qy        319 ATQLQ----KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEA--------- 365
                |     |||  |     :| : :: |   |   : |  :   :   |
Db       3845 VNDLSVEKGKNLILECTYTGTPPISVTWK---KNGVILKHSEKCSITTTETSAILEIPNS 3901

Qy        366 -----GMWQC-LLSDSGQVLLESNIKVL--PTWSTPVHP 396
                  | : | : :||||       | :|   | : ||: |
Db       3902 KLEDQGQYSCHIENDSGQDNCHGAITILEPPYFVTPLEP 3940


RESULT 89
O88654
ID   O88654      PRELIMINARY;      PRT;   340 AA.
```

```
AC   O88654;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NK cell receptor 2B4 splice variant.
GN   NMRK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RA   Stepp S.E., Schatzle J.D., Bennett M., Kumar V., Mathew P.A.;
RT   "Characterization of genomic structure and alternative splicing of the
RT   murine NK cell receptor 2B4.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF082803; AAC34859.1; -.
DR   MGD; MGI:109294; Nmrk.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   SMART; SM00409; IG; 1.
KW   Receptor.
SQ   SEQUENCE   340 AA;  38397 MW;  0D5275A1A354FC1E CRC64;

  Query Match             5.6%;  Score 129.5;  DB 11;  Length 340;
  Best Local Similarity   23.7%;  Pred. No. 0.095;
  Matches   72;  Conservative   50;  Mismatches  121;  Indels   61;  Gaps   11;

Qy         18 LLPAATQG------NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQ-----IKILGNQG 66
              ||  | ||      :: |:|  |  |:|  :  | | :   || : |     |:||
Db         12 LLLRAHQGQDCPDSSEEVVGVSGKPVQLRPSNIQTKDVSVQWKKTEQGSHRKIEIL---- 67

Qy         67 SFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEV-----QLL 121
              ::   |||  |          :| |:| | ||: |::||  |: |: :   :|     |||
Db         68 NWYNDGPSWSNVSFSDIYG-FDYGDFALSIKSAKLQDSGHYLLEITNTGGKVCNKNFQLL 126

Qy        122 VFGLTANSDTHLLQGQ-------SLTLTLESPPGSSPSVQCRSPRGKNI---QGGKTLSV 171
              :     : :|  |: |       :  | |        :|      ||  :   |   |
Db        127 ILD---HVETPNLKAQWKPWTNGTCQLFLSCLVTKDDNVSYALYRGSTLISNQRNSTHWE 183

Qy        172 SQLELQDSGTWTC-------------------TVLQNQKKVEFKIDIVVLAFQKASSI- 210
              :|::     |:||                 :|  | : : | : ||:|      :|
Db        184 NQIDASSLHTYTCNVSNRASWANHTLNFTHGCQSVPSNFRFLPFGVIIVILVTLFLGAII 243

Qy        211 ---VYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAE---RASSSKSWITFDLKNKEVSVK 264
                 |: |: :|::||     |: :           :   ||| | | :  | : :    :
Db        244 CFCVWTKKRKQLQFSPKEPLTIYEYVKDSRASRDQQGCSRASGSPSAVQEDGRGQRELDR 303

Qy        265 RVTQ 268
              ||::
Db        304 RVSE 307


RESULT 90
Q9D7B8
ID   Q9D7B8      PRELIMINARY;      PRT;   287 AA.
AC   Q9D7B8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   2310016B05Rik protein.
GN   D11ERTD736E OR 2310016B05RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Tongue;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK009375; BAB26251.1; -.
DR   MGD; MGI:1289168; D11Ertd736e.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 1.
DR   PROSITE; PS50835; IG_LIKE; 1.
SQ   SEQUENCE   287 AA;  32061 MW;  D9324D3308A03639 CRC64;

  Query Match             5.5%;  Score 129;  DB 11;  Length 287;
  Best Local Similarity   25.2%;  Pred. No. 0.082;
  Matches   54;  Conservative   36;  Mismatches   84;  Indels   40;  Gaps    9;

Qy          8 RHLLLVLQLALLPA--ATQGNKVVLGKKGDTVELTCTASQK-KSIQFHWKNSNQIK---- 60
              | |:|:    :||   | :| | : | :|||| | ||  :| |  : :|     |
Db          2 RPLVLLWGCLVLPGYEALKGPKEISGFEGDTVSLRCTYVEKMKEHRKYWCRQGGILVSRC 61

Qy         61 ---ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVE----D 113
                 :  ||   :|:|   :    || : |    :  :|:::| ::||  | | ::    |
Db         62 GDIVYANQDQEVTRGRMSIR---DSPQEL----SMTVIMRDLTLKDSGKYWCGIDRLGRD 114

Qy        114 QKEEVQLLVF----------GLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSP----R 159
               :  ||  |:||           ||   |:     :::  ::  |  | | |:  :|
Db        115 ESFEVTLIVFPGSYRPVVWLPLTTPQDS-----RAVASSVSKPSVSIPMVRMMAPVLILL 169

Qy        160 GKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
               :  |     | :     |  |  |  :||
Db        170 SLLLAAGLIAFGSHMLRWRKKAWLATETQKNEKV 203


RESULT 91
Q7ZU00
ID   Q7ZU00      PRELIMINARY;      PRT;   564 AA.
AC   Q7ZU00;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Neurolin.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Body;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
```

```
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Body;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC050482; AAH50482.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003596; Ig_v.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 1.
DR   SMART; SM00406; IGv; 1.
DR   PROSITE; PS50835; IG_LIKE; 3.
DR   PROSITE; PS00290; IG_MHC; 1.
SQ   SEQUENCE   564 AA;  61336 MW;  4E4EE408E05CA985 CRC64;

  Query Match             5.5%;  Score 128.5;  DB 13;  Length 564;
  Best Local Similarity   20.6%;  Pred. No. 0.23;
  Matches   92;  Conservative   61;  Mismatches  137;  Indels  157;  Gaps   21;

Qy          9 HLLLVLQLALLPAA--TQGN--KVVLGKKGDTVELTCTASQKKS---IQFHWKNSNQIKI 61
              | :: |  | : ||    |:    |:|  |:|:|: |     |    |   ||    :|
Db          2 HSVICLFGAFIAAAVFAPGSCLPTVIGLYGETIEVPCNNGNNKPDGLIFTKWK---YVKD 58

Qy         62 LGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLL 121
               |: |  | |  |                        |  :  ::  |  |
Db         59 DGSPGDLLIKQAQK--------------------------DDPTVSAMDGYKTRV--- 87

Qy        122 VFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGT 181
                 : |||                                    :| ::|  | |
Db         88 --SIAANS-----------------------------------SLLIAQGSLTDQRV 107

Qy        182 WTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQ 241
              :|| |: :    || ::: |   :| |: | | : :::|         |||  ||   :
Db        108 FTCMVVSSTNLEEFSVEVKV--HKKPSAPVIKNKVKELENG--------KLTQLGECVVE 157

Qy        242 AERASSSKSWITFDLKNKEVSV---------KRVTQDP---------------KLQMGK 276
              :   ::   |    :|| :  |           ||:||               |  :
Db        158 SANPAADLIW----MKNNQALVDDGKTIIITSDVTKDPVTGLSSTSSRLQYTARKEDVAS 213

Qy        277 KLPL---HLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQK----NLTCE 329
              : |    |:| |  :     |   |  :   |    ::|:| |:  : :::     | |:
Db        214 QFPCVAKHVTGPNQV-----STPDTFQIRYPT----EKVSLQVVSQSPIREGDDVTLKCQ 264

Qy        330 VWGPTSPKLMLSLKLENKEAKVSKREKPVWVL----NPEAGMWQCLLSDSGQVLLESNIK 385
              |   |    |     |  ||:  :| |: |      ::|:::| | |:   ::||
Db        265 ADGNPPP---TSFNFNIKGKKVTVTDKDVYTLTGVTRADSGVYKCSLLDND--VMESTQI 319

Qy        386 VLPTWSTPVHPRASALPA--PPTGSAL 410
              |          | | |   |||  |
Db        320 V----------TVSFLDASLTPTGKVL 336


RESULT 92
Q9W213
ID   Q9W213      PRELIMINARY;      PRT;  1395 AA.
AC   Q9W213;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG13521 protein.
GN   ROBO OR CG13521.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Amanatides P.G., Brandon R.C., Rogers Y., An H., Baldwin D.,
RA   Banzon J., Beeson K.Y., Busam D.A., Carlson J.W., Center A.,
RA   Champe M., Davenport L.B., Dietz S.M., Dodson K., Dorsett V.,
RA   Doup L.E., Doyle C., Dresnek D., Farfan D., Ferriera S., Frise E.,
RA   Galle R.F., Garg N.S., George R.A., Gonzalez M., Houck J.,
RA   Hoskins R.A., Hostin D., Howland T.J., Ibegwam C., Jalali M.,
RA   Kruse D., Li P., Mattei B., Moshrefi A., McIntosh T.C., Moy M.,
RA   Murphy B., Nelson C., Nelson K.A., Nunoo J., Pacleb J., Paragas V.,
RA   Park S., Patel S., Pfeiffer B., Phouanenavong S., Pittman G.S.,
RA   Puri V., Richards S., Scheeler F., Stapleton M., Strong R.,
RA   Svirskas R., Tector C., Tyler D., Williams S.M., Zaveri J.S.,
RA   Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
```

```
RA   Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA   Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
RA   Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
RA   Whitfield E.J., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003458; AAM71113.1; -.
DR   HSSP; P56276; 1TLK.
DR   FlyBase; FBgn0005631; robo.
DR   GO; GO:0005886; C:plasma membrane; IDA.
DR   GO; GO:0007411; P:axon guidance; IMP.
DR   InterPro; IPR003962; FnIII_subd.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF00047; ig; 5.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PROSITE; PS50835; IG_LIKE; 5.
KW   Repeat.
SQ   SEQUENCE   1395 AA;  151759 MW;  25CED7DEB44F13F0 CRC64;

  Query Match             5.5%;  Score 128.5;  DB 5;  Length 1395;
  Best Local Similarity   20.1%;  Pred. No. 0.81;
  Matches   97;  Conservative   58;  Mismatches  179;  Indels  149;  Gaps   18;

Qy         26 NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS 85
              || | |  |  |:| | ||        |       ::      |       : :  ||
Db        353 NKKV-GLNG-VVQLPCMASGNPPPSVFWTKEGVSTLM-----FPNSSHGRQHVAADG--- 402

Qy         86 LWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTH---LLQGQSLTLT 142
                  | | ::: ||   |:|      :   : ||   :: |     ::|       |
Db        403 -------TLQITDVRQEDEGYYVCSAFSVVDSSTVRVFLQVSSLDERPPPIIQIGPANQT 455

Qy        143 LESPPGSSPSVQCR-----SPRGK-------------NIQGGKTLSVSQLELQDSGTWT 183
              |  | ||  :: ||      ||| |             :|  | :| |  |:| ||||:|
Db        456 L--PKGSVATLPCRATGNPSPRIKWWFHDGHAVQAGNRYSIIQGSSLRVDDLQLSDSGTYT 513

Qy        184 CTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAE 243
              ||                                ||: | |:    ||||  ||  |   |:
Db        514 CTA---------------------------SGERGETSWAATLTVEK-PGSTSLHRAAD 544

Qy        244 ----------------RASSSKSW-------------------ITFDLKNKEVSVKRV 266
                             | | |  |                    : ||:   :   :
Db        545 PSTYPAPPGTPKVLNVSRTSISLRWAKSQEKPGAVGPIIGYTVEYFSPDLQTGWIVAAQR 604

Qy        267 TQDPKLQMGKKLP-------------LHLTLPQALP----------QYAGSGNLTLALEA 303
                | ::  :  |              :::|  |              | : :|: |
Db        605 VGDTQVTISGLTPGTSYVFLVRAENTQGISVPSGLSNTIKTIEADFDAASANDLSAARTL 664

Qy        304 KTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNP 363
               |||  : ::   : |: ::      |       | :  |::  |:| |           |
Db        665 LTGKSVELIDASAINASAVRLEWMLHV--SADEKYVEGLRIHYKDASV----------P 711

Qy        364 EAGMWQCLLSD-SGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDP 422
               |      : | | :  :  |:|    :   : |    :   |: |     :||     |
Db        712 SAQYHSITVMDASASESFVVGNLKKYTKYEFFLTPFFETIEGQPSNS-----KTALTYEDV 766

Qy        423 PAA 425
              |:|
Db        767 PSA 769


RESULT 93
Q8CGB3
ID   Q8CGB3      PRELIMINARY;      PRT;  1413 AA.
AC   Q8CGB3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to uveal autoantigen with coiled-coil domains and ankyrin
DE   repeats.
GN   2700059D02RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Salivary gland;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC042415; AAH42415.1; -.
DR   MGD; MGI:1919815; 2700059D02Rik.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000727; T_SNARE.
DR   Pfam; PF00023; ank; 5.
DR   SMART; SM00248; ANK; 6.
DR   PROSITE; PS50088; ANK_REPEAT; 5.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50192; T_SNARE; 1.
SQ   SEQUENCE   1413 AA;  160917 MW;  2DE27904AF684120 CRC64;

  Query Match             5.5%;  Score 128.5;  DB 11;  Length 1413;
  Best Local Similarity   20.5%;  Pred. No. 0.83;
  Matches   76;  Conservative   72;  Mismatches  144;  Indels   79;  Gaps   12;

Qy          4 GVP----FRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQI 59
              |:|     | :|  |:|:|    :      :| |: :|:     ::::  ::|
Db        419 GIPPHMHSRSMLRPLELSLPSQTSYSENEILKKELETLRTYYDSAKQDRLKFQ------- 471

Qy         60 KILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQ 119
                          ::|  :    ::|       |  : :| ||||  | ::||  ::||
Db        472 --------------NELAHKVAECKAL------ALECERVK-EDSDEQIKQLEDALKDVQ 510

Qy        120 LLVF---GLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQG---GKTLSVSQ 173
                ::  |     || |  :   || |:  ||   ::    : |:::|   | :  | :
Db        511 KRMYESEGKVKQMQTHFLALKE-HLTNEAATGSHRIIEELREQLKDLKGKYEGASAEVGK 569

Qy        174 LELQ-------------DSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVE 220
              |  |               | |     :::  |::: :     :  ::    | : ||:  |
Db        570 LRSQIKQSEMLVGEFKRDEG----RLVEENKRLQKECGTCEVELERRGRRVVELEGQLKE 625

Qy        221 FSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPL 280
                || :|              |:  | || :: |:  |   :  | :| :   |:   |
Db        626 LGAKLALSV-----------PTEKFESMKSSLSNDINEKVKRLAEVGRDYESAQGEIRQL 674

Qy        281 HLTLPQALPQY---AGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPK 337
                 |      |:         |   || |:|: ::|: : ::   |||::          |
Db        675 KRDLESVRAQHIRPEEHEQLRSRLEQKSGELGKKVSELTLKNQTLQKDV---------EK 725

Qy        338 LMLSLKLENKE 348
              |    || |::
Db        726 LHADNKLLNQQ 736


RESULT 94
O08653
ID   O08653      PRELIMINARY;      PRT;  2629 AA.
AC   O08653;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Telomerase protein component 1.
GN   TLP1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Fischer 344;
RX   MEDLINE=97236507; PubMed=9118230;
RA   Nakayama J., Saito M., Nakamura H., Matsuura A., Ishikawa F.;
RT   "TLP1: a gene encoding a protein component of mammalian telomerase is
```

```
RT   a novel member of WD repeats family.";
RL   Cell 88:875-884(1997).
DR   EMBL; U89282; AAB51690.1; -.
DR   PIR; T32735; T32735.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   InterPro; IPR008850; TEP1_N.
DR   InterPro; IPR008858; TROVE.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF05729; NACHT; 1.
DR   Pfam; PF05386; TEP1_N; 4.
DR   Pfam; PF05731; TROVE; 1.
DR   Pfam; PF00400; WD40; 15.
DR   ProDom; PD000018; WD40; 1.
DR   SMART; SM00320; WD40; 15.
DR   PROSITE; PS50837; NACHT; 1.
DR   PROSITE; PS00678; WD_REPEATS_1; 1.
DR   PROSITE; PS50082; WD_REPEATS_2; 7.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW   Repeat; WD repeat.
SQ   SEQUENCE   2629 AA;  291705 MW;  B90C3F72E5D4B6BE CRC64;

  Query Match             5.5%;  Score 127;  DB 11;  Length 2629;
  Best Local Similarity   20.3%;  Pred. No. 2.6;
  Matches   88;  Conservative   62;  Mismatches  133;  Indels  150;  Gaps   20;

Qy          9 HLLLVLQLALL---------PAATQGNKVVLGKKGDTVEL---------TCTASQKKSI- 49
              | ||| |:  |           ||:: : ::|     :| |          ||       |
Db       2235 HPLLVCQIHTLQGHSGPVTAAAASEASGLLLTSDNSSVRLWQIPKEADDTCKPRSSAVIT 2294

Qy         50 QFHWKNSNQIKILGNQGSFLT-----------KGPSKLND------------RADSRRSL 86
                 |      : : ||:   ||           : | :::|              |:   |
Db       2295 AVAWAPDGSLVVSGNEAGELTLWQKAQAVATARAPGRVSDLIWCSANAFFVLSANENVSE 2354

Qy         87 WD----QG----NFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQS 138
              |      :|    || | :| :  ||                  :: |:     |     |||
Db       2355 WQVELRKGSTCTNFRLYLKRVLQEDLG---------------VLTGMALAPD-----GQS 2394

Qy        139 LTL------TLESPPGSSPSVQCRSPRGKN-------------IQGGKTLSVSQLELQDS 179
              | |      |:  |||:|| ||     :              :| | : |:| || ::|
Db       2395 LILMKEDVELLQMKPGSTPSSICRRYAVHSSILCTSKDYGLFYLQQGNSGSLSILEQEES 2454

Qy        180 GTWTCTVLQNQKKVEFK----------IDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTV 229
              | :       :| ::|           : |     :  ||::          |  : : :
Db       2455 GKF-------EKTLDFNLNLNNPNGSPVSITQAEPESGSSLL-------CATSDGMLWNL 2500

Qy        230 EKLTGSGE-----LWWQAERASSSK-------------SWITFDLKNKEVSVKRVTQDP 270
               : |  ||     :| : |   |:              ||:       | :  :  |
Db       2501 SECTPEGEWVVDNIWQKKSRNPKSRTPGTDSSPGLFCMDSWV-------EPTHLKARQCK 2553

Qy        271 KLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTC-E 329
              |: :|    ||:     ||     | |  | | :  || :  :: ::   : :  ::| |
Db       2554 KIHLGSVTALHV-----LP-----GLLVTASEDRDVKLWERPSMQLLGLFRCEGPVSCLE 2603

Qy        330 VWGPTSPKLMLSL 342
              |   |  | |::
Db       2604 PWMEPSSPLQLAV 2616


RESULT 95
Q9ESE5
ID   Q9ESE5      PRELIMINARY;      PRT;   399 AA.
AC   Q9ESE5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NK cell receptor 2B4.
GN   NMRK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=20395298; PubMed=10941850;
RA   Kumaresan P.R., Huynh V.T., Mathew P.A.;
RT   "Polymorphism in the 2B4 gene of inbred mouse strains.";
RL   Immunogenetics 51:758-761(2000).
DR   EMBL; AF184208; AAG12323.1; -.
DR   MGD; MGI:109294; Nmrk.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   SMART; SM00409; IG; 1.
KW   Receptor.
SQ   SEQUENCE   399 AA;  44991 MW;  632E61217FF1D065 CRC64;

  Query Match             5.4%;  Score 126.5;  DB 11;  Length 399;
  Best Local Similarity   21.3%;  Pred. No. 0.2;
  Matches   84;  Conservative   61;  Mismatches  151;  Indels   99;  Gaps   16;

Qy        18 LLPAATQGN------KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTK 71
             ||  | ||       : |:|  |  | |  :  | | :   ||     |: |:  :  |:
Db        12 LLLRAHQGQDCSDSPEEVVGVSGKPVRLRPSNIQTKDVSVQWKK----KVQGSHTN--TE 65

Qy        72 GPSKLNDRADSRRSL--------WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEV----- 118
                |: | || |          :| |:| | ||: :::||  |: |:  :    |
Db        66 ILISCNNNA-SRCSTVVSSDIYGFDYGDFALSIKSAQLQDSGHYLLELTYRNGTVCTKNF 124

Qy       119 QLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----GGKTLSVS-- 172
             |:|:     : :|  |: |     |  :       : |  |:  |:      | || :|
Db       125 QILILD---HVETPHLKAQWNAWTNGT---CQLFLSCSVPKDDNVSYALYRGSTLMISNQ 178

Qy       173 --------QLELQDSGTWTCTVLQ--------------------NQKKVEFKIDIVVLAF 204
                     | :      |:|| |                        | : : | : ||:|
Db       179 RNSTHWENQTDASSLHTYTCNVSNRASWANHTLNFTHGCQSVHLNFRFLPFGVIIVILVT 238

Qy       205 QKASSI----VYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAE---RASSSKSWITFDLK 257
                :|     |: |: :|::||     |: :          :   ||| | | :  | :
Db       239 LFLGAIICFCVWTKKRKQLQFSPKEPLTIYEYVKDSRASRDQQGCSRASGSPSDVQEDGR 298

Qy       258 NKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVM 317
               :    :||::                : : |||     |:          | ::  :
Db       299 GQRELDRRVSE---------------VLEQLPQQNSPGD--------RGTMYSMIQCKPF 335

Qy       318 RATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVS 352
               :|  :|   | |:     |       |  |: : :|
Db       336 DSTSQEK---CTVYSVVQPSRKSGSKKRNQNSSLS 367


RESULT 96
Q26475
ID   Q26475      PRELIMINARY;      PRT;   484 AA.
AC   Q26475;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   REGA-1 protein precursor.
GN   REGA-1.
OS   Schistocerca americana (American grasshopper).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridomorpha;
OC   Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca.
OX   NCBI_TaxID=7009;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96189346; PubMed=8625808;
RA   Seaver E.C., Carpenter E.M., Bastiani M.J.;
RT   "REGA-1 is a GPI-linked member of the immunoglobulin superfamily
RT   present on restricted regions of sheath cell processes in
RT   grasshopper.";
RL   Development 122:567-578(1996).
DR   EMBL; X93601; CAA63800.1; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
```

```
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 2.
DR   PROSITE; PS50835; IG_LIKE; 3.
KW   Immunoglobulin domain; Signal.
FT   SIGNAL       1     21       POTENTIAL.
SQ   SEQUENCE   484 AA;  54013 MW;  0686F9D7D7BEAF9C CRC64;

  Query Match             5.4%;  Score 126.5;  DB 5;  Length 484;
  Best Local Similarity   19.2%;  Pred. No. 0.26;
  Matches   92;  Conservative   74;  Mismatches  201;  Indels  111;  Gaps   18;

Qy         20 PAATQGNKVVLGKKGDTVELTCT-------ASQKKSIQFHWKNSNQIKILGNQGSFLTKG 72
              |  |  : ::   :||||   |:         :: :    |||     |:   |: :
Db         60 PEFTVQSMMLEAHEGDTVTFPCSWKVPVNDLKERPNFPVMWKNGTSSLIM---GAIIATT 116

Qy         73 PSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL----------LV 122
                ::     :            ||: :::  |  || ||:   :  : |          |
Db        117 DKRVQSNGSN-----------LILTDVRPTDRGTYTCEILSSETPMDLHHSLEVFEAPKV 165

Qy        123 FGLTANSDTHLLQ-GQSLTLTLESPPGSSPSVQCRSPRGKNIQ--GGKTLSVSQLELQDS 179
              :  |  : ::: |::|||: ::      |::      |     :    | ||:  ::    |
Db        166 AKIHPNKSSIVVEAGEALTLSCDANGYPKPTISWIKKRKAKEETFKGPTLNFEKVAPSHS 225

Qy        180 GTWTCTVLQNQKK---------VEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE 230
              | : ||      |         | :|  :  :  :  | : |: | |    | | |
Db        226 GEYECTAENTYGKSNPAHVKVVVHYKPKVTAIKERVISGVEYESELECRVDSEPKA---- 281

Qy        231 KLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQ 290
                  :: || :       |          :  |  :    ||:: |             :
Db        282 ------KVKWQRDGRLIEPS--------NDFEVTHMNHIYKLKIRK------------TK 315

Qy        291 YAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL------------TCEVWGPTSPK 337
              :  |  |   |    |      ::|  | :|   | :             | | : | |
Db        316 TSDFGIYTCEAENTLGSAKSTIDLTRMPSTPALKKVNHADDGSSVEFEWTLESYAPIS-- 373

Qy        338 LMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKVLPT---WSTPV 394
                |: || ::  | | |  | : |  ::       | |    |  : :: |    |:
Db        374 -QYELRYRNKMSEDWKSESPA-VTDAEGNIYTVKHKLSPQPPGEYEVVLVSTNEFGSSRS 431

Qy        395 HPRASALP-----APPTGSALPDPQTASALPDP--------PAASALPAALAVISFLL 439
                |    |      | | |   |:|    :|:         |   ||   |:: :| |
Db        432 APHTFQLTHELGVAEPRGE--PEPQ---KIPESGNNGSETRPRLMALLTVLSLSTFFL 484


RESULT 97
Q7TNP4
ID   Q7TNP4      PRELIMINARY;      PRT;   785 AA.
AC   Q7TNP4;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Strausberg R.;
RL   Submitted (AUG-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC056205; AAH56205.1; -.
KW   Hypothetical protein.
FT   NON_TER      1      1
SQ   SEQUENCE   785 AA;  85615 MW;  0B7DB89BDC6D0BA7 CRC64;

  Query Match             5.4%;  Score 126;  DB 11;  Length 785;
  Best Local Similarity   20.5%;  Pred. No. 0.57;
  Matches   80;  Conservative   49;  Mismatches  133;  Indels  128;  Gaps   18;

Qy         18 LLPAATQGNKVVLGKKGDTVELTCTASQKKSI-------QFHW--KNSNQIKILG----- 63
              :||||          :  |  |  | |  :::          | |     |::: :
Db        332 VLPAA---------RPSDGGEFQCVAGDERAYFTVTITDVFSWIVYPSSEVHVAAVRLER 382

Qy         64 ----------------NQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDT 106
                             |  : : |: | :: |: |          |:: ::::|||
Db        383 VVLTCELCRPWAEVRWTKDGEEVVESPALLLEKEDTIRR--------LVLPSVQLEDSGE 434

Qy        107 YICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESPPG----------SSPSVQC 155
              |:||: |                      :  | |:|:  ||              : |:|
Db        435 YLCEIHD--------------------ESASFTITVTEPPVRIIYPQDEVTLHAVSLEC 473

Qy        156 RSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVV----LAFQKASSIV 211
                        :   :|  :| | : |          | : :      :    | |:
Db        474 -----------VVLTCELSREDGGEFVCDA--GDDSAFFTVTVTAPPERIVHPAARSL- 518

Qy        212 YKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPK 271
                      ::|  |          |  ||  :   | |   |    |   || |
Db        519 ------DLQFGAP---------GHVELRCEVAPAGSQVCWYKDGL---EVEV-----SDA 555

Qy        272 LQMGKKLPLH-LTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEV 330
              ||:| : |   ||||  |  |: ||          || |  :       |   | :   | | |
Db        556 LQLGAEGPARTLTLPHAQPEDAGEYVCETRDEAVTFNVSLAELPVQFLAPEAAPNPLCVV 615

Qy        331 WGPTSPKLMLSLKLENKEAKV--SKREKPV 358
                |  | ::|| :|    |:|  |    ||
Db        616 --PGEP-VVLSCELSRASAQVFWSHNGSPV 642


RESULT 98
Q9W1D5
ID   Q9W1D5      PRELIMINARY;      PRT;  3197 AA.
AC   Q9W1D5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG30171 protein.
GN   BEST:HL01080 OR CG3901 OR CG18019 OR CG18020 OR CG18021 OR CG30171.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
```

```
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003462; AAF47134.2; -.
DR   FlyBase; FBgn0040499; BEST:HL01080.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR   GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00047; ig; 15.
DR   Pfam; PF00069; pkinase; 2.
DR   ProDom; PD000001; Prot_kinase; 2.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00409; IG; 21.
DR   SMART; SM00408; IGc2; 18.
DR   SMART; SM00220; S_TKc; 2.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50835; IG_LIKE; 14.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 2.
KW   ATP-binding; Immunoglobulin domain; Transferase.
SQ   SEQUENCE   3197 AA;  361286 MW;  F4E74688B455A3B6 CRC64;

  Query Match             5.4%;  Score 126;  DB 5;  Length 3197;
  Best Local Similarity   19.6%;  Pred. No. 4;
  Matches   91;  Conservative   74;  Mismatches  154;  Indels  146;  Gaps   25;

Qy         22 ATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRAD 81
              |: |   :|        || |     | ::|  ||:  ::  : ::  |: :
Db        303 ASIGGSAIL-------ELQCKGFPKPAVQ--WKHDGEVIQVDDRHKFMYE---------- 343

Qy         82 SRRSLWDQGNFPLIIKNLKIEDSDTY----ICEVEDQKEEVQLLVFG---LTANSDTHLL 134
                    |: :  |:|||:   |:  |     | |:   :  : |:|     :   :|
Db        344 ------DEESMSLVIKNVDTVDAGVYTIEAINELGQDESSINLVVKAPPKIKKITDITCS 397

Qy        135 QGQSLTLTLESPPGSSPSVQCRSPRGKNIQG----------GKTL-----SVSQLELQD 178
              |::: : :|      |:||  :  ||::             ||:|     | :::|
Db        398 AGETIKMEIEVEGFPQPTVQVTN-NGKDVTAESNVKISSSSIGKSLEKVVVEVKEIKLSQ 456

Qy        179 SGT--------------WTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSF 223
              :|               |:|||             ::|  |:   |     | | |: :
Db        457 AGNYSIKATNDLSQTSEYWSCTVKSK--------PVIVKNFE---SEYIHGEKENVQMTV 505

Qy        224 PL-AFTVEKLTGSGELWWQAE---RASSSKSWITFD-----LK---NKEVSVKRVTQDPK 271
              : |:   |||      |:  |   : : ||  :: |     ||       |   : |
Db        506 RIDAYPEAKLT-----WYHDETEIKITDSKYTVSSDGNAYTLKITGATRVDAGKYTVKAT 560

Qy        272 LQMGKKLPLHLTLPQALPQYA----------GSGNLTLAL-------------------- 301
              : |        | :  |::           |  |: | :
Db        561 NEHGSATSSTQLLIKCAPEFTHKLKNITVAEGDSNVELVVGVDAYPRPHAKWYIDGIEID 620

Qy        302 EAKTGKLHQE----VNLVVMR-ATQLQKNLTCEVWGP-------------TSPKLMLSLK 343
              | :    | |       |:: : || :| | ||::                   ||:   ||
Db        621 EKRNDFRHVEEGNDFKLIMNQVATNMQGNYTCKIMNDYGKLEDNCVVTVNCKPKVKRGLK 680

Qy        344 -LENKEAKVSKREKPVWVLNPEAGM-WQCLLSDSGQVLLESNIKV 386
              :| :| |    |  |:   ||| : |     |  ::  :: ||:
Db        681 NVEVQEGKSFTLEVEVY-SEPEAKIKW---FKDGHEIYEDARIKI 721


RESULT 99
O00481
ID   O00481      PRELIMINARY;      PRT;   513 AA.
AC   O00481;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Butyrophilin.
GN   BTF5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ruddy D.A., Kronmal G.S., Lee V.K., Mintier G.A., Quintana L.,
RA   Domingo R. Jr., Meyer N.C., Basava A., McClelland E., Fullan A.,
RA   Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihashi Z.,
RA   Wolff R.K., Schatzman R.C., Feder J.N.;
RT   "A 1.1 megabase transcript map of the human hereditary hemochromatosis
RT   locus.";
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U90552; AAB53430.1; -.
DR   Genew; HGNC:1138; BTN3A1.
DR   GO; GO:0006629; P:lipid metabolism; TAS.
DR   InterPro; IPR001870; B302.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR006574; PRY.
DR   InterPro; IPR003877; SPRY_receptor.
DR   Pfam; PF00622; SPRY; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00589; PRY; 1.
DR   SMART; SM00449; SPRY; 1.
DR   PROSITE; PS50835; IG_LIKE; 2.
SQ   SEQUENCE   513 AA;  57762 MW;  CD334D7727CD1F63 CRC64;

  Query Match             5.4%;  Score 125.5;  DB 4;  Length 513;
  Best Local Similarity   21.3%;  Pred. No. 0.34;
  Matches  102;  Conservative   72;  Mismatches  191;  Indels  113;  Gaps   25;

Qy          7 FRHLLLVLQLALLPAATQGNKV-----VLGKKGDTVELTC---TASQKKSIQFHWKNSNQ 58
              ||  ||:||| |:| : | : :     :|   |:  :| |       '::::  | :|:
Db         14 FRVCLLLLQL-LMPHSAQFSVLGPSGPILAMVGEDADLPCHLFPTMSAETMELKWVSSSL 72

Qy         59 IKIL-----GNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED 113
               :::     | :       | :   |     |      |   | | |:   ||  |:|  :|
Db         73 RQVVNVYADGKEVEDRQSAPYR--GRTSILRDGITAGKAALRIHNVTASDSGKYLCYFQD 130

Qy        114 ----QKEEVQLLVFGLTANSDTHL----LQGQSLTLTLESPPG-SSPSVQCRSPRGKNIQ 164
                  :|  |:| |  |   || |:     :  : |    |      | :|  : :|:||
Db        131 GDFYEKALVELKVAAL--GSDLHVDVKGYKDGGIHLECRSTGWYPQPQIQWSNNKGENIP 188

Qy        165 -----------GGKTLSVSQLELQDSGTW-TCTVLQNQKKVEFKIDIVVLAFQKASSI-- 210
                        |   :: | :    ||   :||:  :           :|  :| :||
Db        189 TVEAPVVADGVGLYAVAASVIMRGSSGEGVSCTIRSS----------LLGLEKTASISI 237

Qy        211 --VYKKEGEQ--VEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
                 : :  ::     :  |   :  | |:|   || :    ::    |  | :|  :: :
Db        238 ADPFFRSAQRWIAALARTLPVLLLLLGGAGYFLWQQQEEKKTQ----FRKKKREQELREM 293

Qy        267 T-QDPKLQMGKKLPLHLTLPQALPQYAGSGN--------------LTLALEAKTGK--- 307
                | :   :: |   |     |||  |                 : |: ||
Db        294 AWSTMKQEQSTRVKLLEELRWRSIQYASRGERHSAYNEWKKALFKPADVILDPKTANPIL 353

Qy        308 LHQEVNLVVMRATQLQK-------------NLTCEVWGPTSPKLMLSLKL-ENKEAKVSK 353
              |  |     | || : |                | || :   |  :    ::: : ||  :
Db        354 LVSEDQRSVQRAKEPQDLPDNPERFNWHYCVLGCESF--ISGRHYWEVEVGDRKEWHIGV 411

Qy        354 REKPV----WV-LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRAS-ALPAPP 405
                | |     || : || | |    |:|  :            : |   || :  || ||
Db        412 CSKNVQRKGWVKMTPENGFWTMGLTDGNK------------YRTLTEPRTNLKLPKPP 457


RESULT 100
Q921P2
ID   Q921P2      PRELIMINARY;      PRT;   605 AA.
AC   Q921P2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   NCAM1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011310; AAH11310.1; -.
DR   MGD; MGI:97281; Ncam1.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig_c2.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 5.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00408; IGc2; 5.
DR   PROSITE; PS50835; IG_LIKE; 5.
KW   Hypothetical protein; Immunoglobulin domain.
SQ   SEQUENCE   605 AA;  67353 MW;  DBD98C64D245A867 CRC64;

  Query Match             5.4%;  Score 125;  DB 11;  Length 605;
  Best Local Similarity   19.9%;  Pred. No. 0.47;
  Matches   61;  Conservative   48;  Mismatches  139;  Indels   58;  Gaps   12;

Qy          3 RG-VPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHW-KNSNQIK 60
              || : |: : ::: :     | |          | :| | | |          | |:   |:
Db        196 RGEINFKDIQVIVNVPPTVQARQSIVNATANLGQSVTLVCDADGFPEPTMSWTKDGEPIE 255

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVE----DQKE 116
                          |:  |  : ::   :  | |:|:   |   |:|  |    :|
Db        256 ----------------NEEEDDEKHIFSDDSSELTIRNVDKNDEAEYVCIAENKAGEQDA 299

Qy        117 EVQLLVFG----LTANSDTHLLQGQSLTLTLESPPGSSPSVQCR------SPRGKNIQGG 166
              : | ||         : | :   : :||| |:     ||:  |      |   | : |
Db        300 SIHLKVFAKPKITYVENQTAMELEEQVTLTCEASGDPIPSITWRTSTRNISSEEKTLDGH 359

Qy        167 ---------KTLSVSQLELQDSGTWTCTVLQ--NQKKVEFKIDIVVLAFQKASSIVYKKE 215
                      :|::  ::  |:| : ||      |      :::      :   || |
Db        360 MVVRSHARVSSLTLKSIQYTDAGEYICTASNTIGQDSQSMYLEVQYAPKLQGPVAVYTWE 419

Qy        216 GEQVE-----FSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKN-KEVSVKRVTQD 269
              | ||       |::| | |:          |::  :   | ::    : |    |   || |
Db        420 GNQVNITCEVFAYPSA-TIS--------WFRDGQLLPSSNYSNIKIYNTPSASYLEVTPD 470

Qy        270 PKLQMG 275
               :   |
Db        471 SENDFG 476


Search completed: August  3, 2004, 13:13:35
Job time : 38.1057 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:09 ; Search time 46.0872 Seconds
(without alignments)
2777.216 Million cell updates/sec

Title: SEQ5
Perfect score: 2325
Sequence: 1 MNRGVPFRHLLLVLQLALLP.....VISFLLGLGLGVACVLARTR 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2049	88.1	474	3	AAY59170	Aay59170 CD4-Ig fu
2	2047.5	88.1	436	3	AAY51080	Aay51080 Human fus
3	2046	88.0	458	3	AAB07769	Aab07769 DNA encod
4	2044.5	87.9	729	3	AAB19507	Aab19507 CD4-IgG1
5	2042	87.8	729	1	AAP93008	Aap93008 Genetic c
6	2042	87.8	729	3	AAY51078	Aay51078 Human fus
7	2042	87.8	729	3	AAY59168	Aay59168 CD4-Ig fu
8	2040.5	87.8	416	3	AAB19509	Aab19509 CD4-IgM f
9	2040	87.7	462	2	AAR27277	Aar27277 CD4:eta p
10	2040	87.7	462	2	AAR78677	Aar78677 T-cell re
11	2040	87.7	462	2	AAR89457	Aar89457 CD4:gamma
12	2040	87.7	462	2	AAW02214	Aaw02214 CD4:Fc re
13	2040	87.7	462	2	AAW83142	Aaw83142 Chimeric
14	2038.5	87.7	575	2	AAR27276	Aar27276 CD4:zeta
15	2038.5	87.7	575	2	AAR78676	Aar78676 T-cell re
16	2038.5	87.7	575	2	AAR89456	Aar89456 CD4:zeta
17	2038.5	87.7	575	2	AAW02213	Aaw02213 CD4:T-cel
18	2038.5	87.7	575	2	AAW83140	Aaw83140 Chimeric
19	2038	87.7	458	1	AAP81990	Aap81990 Clone pT4
20	2038	87.7	458	1	AAP91369	Aap91369 T4 protei
21	2038	87.7	458	2	AAY39826	Aay39826 Soluble h
22	2038	87.7	481	3	AAB19510	Aab19510 CD4-IgM f
23	2038	87.7	481	3	AAY59171	Aay59171 CD4-Ig fu
24	2037	87.6	398	2	AAR89450	Aar89450 CD4 D1-D4
25	2037	87.6	532	2	AAR27278	Aar27278 CD4:gamma
26	2037	87.6	532	2	AAR78678	Aar78678 T-cell re
27	2037	87.6	532	2	AAR89458	Aar89458 CD4:eta f
28	2037	87.6	532	2	AAW02215	Aaw02215 CD4:T-cel
29	2037	87.6	532	2	AAW83141	Aaw83141 Chimeric
30	2037	87.6	616	3	AAY51082	Aay51082 Human fus
31	2037	87.6	616	3	AAY59172	Aay59172 CD4-Ig fu
32	2034.5	87.5	631	1	AAP93009	Aap93009 Genetic c
33	2034.5	87.5	631	3	AAB19508	Aab19508 CD4-IgG1
34	2034.5	87.5	631	3	AAY51079	Aay51079 Human fus
35	2034.5	87.5	631	3	AAY59169	Aay59169 CD4-Ig fu
36	2034	87.5	398	2	AAR78673	Aar78673 CD4 domai
37	2032	87.4	458	3	AAY88329	Aay88329 T4 glycop
38	2032	87.4	458	4	AAB81502	Aab81502 Human CD4
39	2032	87.4	458	7	ADD25609	Add25609 Binding d
40	2032	87.4	458	7	ADE57489	Ade57489 Human Pro
41	2032	87.4	473	7	ADA44807	Ada44807 CD4/TCR C
42	2032	87.4	519	2	AAR20152	Aar20152 Human CD4
43	2031	87.4	400	2	AAR06374	Aar06374 Truncated
44	2031	87.4	481	3	AAY51081	Aay51081 Human fus
45	2031	87.4	2037	2	AAR04032	Aar04032 Full leng
46	2031	87.4	2050	2	AAR07641	Aar07641 Deduced s
47	2030	87.3	458	2	AAR13491	Aar13491 Human CD4
48	2030	87.3	481	1	AAP93011	Aap93011 Genetic c
49	2026	87.1	614	1	AAP93012	Aap93012 Genetic c
50	2024	87.1	458	4	AAG79087	Aag79087 Amino aci
51	2018	86.8	394	1	AAP93506	Aap93506 Derived s
52	2017	86.8	402	1	AAP91922	Aap91922 Sequence
53	2017	86.8	402	1	AAP94757	Aap94757 Sequence
54	2015	86.7	394	3	AAY88328	Aay88328 T4 glycop
55	2014	86.6	458	7	ADE65841	Ade65841 Human CD4
56	2012	86.5	394	2	AAY39825	Aay39825 Soluble h
57	2010	86.5	458	2	AAR06373	Aar06373 T4 encode
58	2010	86.5	524	1	AAP94703	Aap94703 Sequence
59	2010	86.5	2458	2	AAR07640	Aar07640 Deduced p
60	2010	86.5	2458	2	AAR04031	Aar04031 Full leng
61	2008	86.4	458	2	AAR11285	Aar11285 gp120 bin
62	2008	86.4	458	2	AAR10988	Aar10988 Chimpanze
63	2002	86.1	458	2	AAR04910	Aar04910 T4 protei
64	2001	86.1	394	3	AAB07768	Aab07768 The solub
65	2001	86.1	399	2	AAR20151	Aar20151 Chimpanze
66	1996	85.8	400	2	AAR20150	Aar20150 Chimpanze
67	1990	85.6	399	1	AAP93010	Aap93010 Genetic c
68	1974	84.9	616	3	AAB19511	Aab19511 CD4-IgG1
69	1956	84.1	400	2	AAR20154	Aar20154 Sol. rhes
70	1934	83.2	458	2	AAR15149	Aar15149 CD4 coord
71	1915	82.4	384	1	AAP90833	Aap90833 Amino aci
72	1911	82.2	942	2	AAR41041	Aar41041 CD4-GBP13
73	1911	82.2	1786	2	AAR41043	Aar41043 CD4-EBA17
74	1909	82.1	433	2	AAW41376	Aaw41376 Human CD4
75	1908	82.1	729	2	AAR41042	Aar41042 CD4-GBPH
76	1904	81.9	375	2	AAR07721	Aar07721 Recombina
77	1904	81.9	433	3	AAY54500	Aay54500 Amino aci
78	1904	81.9	434	1	AAP96151	Aap96151 Sequence
79	1904	81.9	434	1	AAP93557	Aap93557 Fusion of
80	1901	81.8	432	2	AAR74222	Aar74222 Epitope o
81	1901	81.8	432	2	AAY30514	Aay30514 Predicted
82	1901	81.8	432	6	ADA25188	Ada25188 CD4 epito
83	1899	81.7	370	1	AAP93528	Aap93528 Human sol
84	1899	81.7	370	4	AAB83356	Aab83356 Human CD4
85	1891	81.3	369	2	AAY39824	Aay39824 Soluble h
86	1891	81.3	369	3	AAY88327	Aay88327 T4 glycop
87	1887	81.2	370	2	AAR12956	Aar12956 Lys(64) C
88	1886	81.1	370	2	AAR12954	Aar12954 His(40) C
89	1886	81.1	370	2	AAR12964	Aar12964 Val(143)
90	1886	81.1	370	2	AAR12965	Aar12965 Arg(150)
91	1885	81.1	370	2	AAR12958	Aar12958 Gln(80) C
92	1882	80.9	370	2	AAR12951	Aar12951 Glu(17),
93	1880	80.9	370	2	AAR12957	Aar12957 Asn(72),
94	1879	80.8	370	2	AAR12959	Aar12959 Asn(88),
95	1879	80.8	370	2	AAR12962	Aar12962 Leu(127),
96	1878	80.8	435	1	AAP90992	Aap90992 Human CD4
97	1874	80.6	370	2	AAR12966	Aar12966 Thr(162),
98	1873	80.6	370	2	AAR12952	Aar12952 Ile(23),

99	1872	80.5	370	2	AAR12967	Aar12967 Asp(155),
100	1871	80.5	435	1	AAP91289	Aap91289 Sequence
101	1870	80.4	370	2	AAR12955	Aar12955 Gly(48),
102	1869	80.4	370	2	AAR12963	Aar12963 His(132),
103	1869	80.4	370	2	AAR12960	Aar12960 Lys(99),
104	1867	80.3	433	2	AAR08335	Aar08335 CD4. 3/20
105	1866	80.3	370	2	AAR12961	Aar12961 Ser(121),
106	1862	80.1	458	2	AAR10987	Aar10987 Rhesus mo
107	1861	80.0	458	2	AAR20148	Aar20148 Cynomolgu
108	1860	80.0	370	2	AAR12953	Aar12953 Thr(27),
109	1859	80.0	400	2	AAR20149	Aar20149 Sol. rhes
110	1368	58.8	318	2	AAR07606	Aar07606 Plasmid T
111	1363	58.6	295	2	AAR07605	Aar07605 Plasmid T
112	1262	54.3	273	2	AAR21078	Aar21078 C-termina
113	1113.5	47.9	454	2	AAR33181	Aar33181 L3T4 muta
114	1080	46.5	474	2	AAR54201	Aar54201 Feline T
115	1059	45.5	507	3	AAB00158	Aab00158 sCD4-SCFv
116	1050.5	45.2	310	2	AAR26784	Aar26784 CD4-IgG2
117	1050.5	45.2	310	2	AAR46680	Aar46680 CD4-kappa
118	1050.5	45.2	310	3	AAY85081	Aay85081 CD4-kappa
119	1050.5	45.2	310	4	AAB67324	Aab67324 CD4-kappa
120	1050.5	45.2	310	4	AAB80885	Aab80885 Human CD4
121	1050.5	45.2	310	6	ABG71124	Abg71124 CD4-kappa
122	1048	45.1	534	2	AAR26531	Aar26531 Sequence
123	1046	45.0	530	2	AAR26783	Aar26783 CD4-IgG2
124	1046	45.0	530	3	AAY85080	Aay85080 CD4-IgG2
125	1046	45.0	530	4	AAB67323	Aab67323 CD4-IgG2

ALIGNMENTS

RESULT 1
AAY59170
ID AAY59170 standard; protein; 474 AA.
XX
AC AAY59170;
XX
DT 14-MAR-2000 (first entry)
XX
DE CD4-Ig fusion protein CD4Mmu.
XX
KW HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
KW secreted protein; SIV infection; medicament.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN CA1340741-C.
XX
PD 14-SEP-1999.
XX
PF 20-JAN-1989; 89CA-00588749.
XX
PR 20-JAN-1989; 89CA-00588749.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-063015/06.
DR N-PSDB; AAZ48203.
XX
PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
PT the treatment of HIV or simian immunodeficiency virus infections.
XX
PS Example 1; Page 47-53; 89pp; English.
XX
CC The invention provides a fusion gene encoding a fusion protein that
CC comprises an extracellular CD4 DNA sequence or its fragment which binds
CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
CC sequence of an Ig heavy or light chain, where the DNA sequence encoding
CC the variable region has been replaced with the DNA sequence which encodes
CC extracellular CD4 or its gp120 binding fragment. The fusion protein is
CC capable of being secreted. The fusion proteins are useful for treating
CC HIV or SIV infections in animals, preferably humans. They are also useful
CC for producing medicaments which can be used for treating HIV or SIV
CC infections in humans. The present sequence represents the fusion protein
CC CD4Mmu where the CD4 is linked to human IgG1 at the Mst2 site upstream of
CC the CH1 region
XX
SQ Sequence 474 AA;

Query Match 88.1%; Score 2049; DB 3; Length 474;
Best Local Similarity 91.6%; Pred. No. 1.1e-131;
Matches 402; Conservative 6; Mismatches 21; Indels 10; Gaps 1;

```
Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV----------HPRASALPAPPTGSAL 410
              ||||||||||||||||||||||||||||||||||          ||:  | :  : :
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPELPLTAHPKGQTLHSLSSDTCP 420

Qy        411 PDPQTASALPDPPAASALP 429
              | |   : |  | | :  |
Db        421 PKPTPKAKLSTPSARTPAP 439
```

RESULT 2
AAY51080
ID AAY51080 standard; protein; 436 AA.
XX
AC AAY51080;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human fusion protein CD4Mg.
XX
KW Fusion protein; human; CD4; IgM; immunoglobulin; gp120;
KW anti-human immunodeficiency virus; CD4Mg.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US6004781-A.
XX
PD 21-DEC-1999.
XX
PF 04-FEB-1994; 94US-00191708.
XX
PR 22-JAN-1988; 88US-00147351.
PR 23-JAN-1989; 89US-00299596.
PR 09-JUN-1992; 92US-00896781.
PR 12-APR-1993; 93US-00057952.

```
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-085792/07.
DR   N-PSDB; AAZ44063.
XX
PT   Fusion protein useful for the treatment of human immunodeficiency virus.
XX
PS   Example 1; Col 41-50; 39pp; English.
XX
CC   This invention describes a novel nucleic acid (I) encoding a fusion
CC   protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
CC   and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
CC   chain (III). The products pf the invention have anti-human
CC   immunodeficiency virus (HIV) activity and are capable of binding to
CC   gp120. The fusion protein is useful for treating human immunodeficiency
CC   virus (HIV) or simian immunodeficiency virus (SIV). This sequence
CC   represents the fusion protein CD4Mg which is constructed from CD4 linked
CC   to human IgM upstream of the CH1 region
XX
SQ   Sequence 436 AA;

  Query Match            88.1%;  Score 2047.5;  DB 3;  Length 436;
  Best Local Similarity  93.0%;  Pred. No. 1.3e-131;
  Matches  401;  Conservative    3;  Mismatches   12;  Indels   15;  Gaps    1;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||:     |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH----------ADPECPXPKP 405

Qy        421 DPPAASALPAA 431
              | |  : |:|
Db        406 TPKAKLSTPSA 416


RESULT 3
AAB07769
ID   AAB07769 standard; protein; 458 AA.
XX
AC   AAB07769;
XX
DT   07-NOV-2000  (first entry)
XX
DE   DNA encoding a human T4 glycoprotein.
XX
KW   Human; T4 glycoprotein; human immunodeficiency virus; HIV;
KW   envelope glycoprotein; AIDS; virus binding.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /note= "leader sequence"
FT   Modified-site   296. .298
FT                   /note= "N-linked glycosylation site"
FT   Modified-site   325. .327
FT                   /note= "N-linked glycosylation site"
FT   Domain          398. .420
FT                   /note= "transmembrane domain"
FT   Domain          421. .458
FT                   /note= "cytoplasmic domain"
XX
PN   US6093539-A.
XX
PD   25-JUL-2000.
XX
PF   06-JUN-1995;   95US-00466368.
XX
PR   21-AUG-1986;   86US-00898587.
PR   11-JUN-1991;   91US-00713564.
PR   06-JUL-1992;   92US-00909021.
PR   12-DEC-1994;   94US-00354452.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Maddon PJ,  Chess L,  Axel R,  Weiss R,  Mcdougal JS,  Littman DR;
XX
DR   WPI; 2000-505203/45.
DR   N-PSDB; AAA59352.
XX
PT   New isolated nucleic acid encoding a human T cell surface protein and the
PT   soluble surface T4 glycoprotein that it encodes, useful as prophylaxis
PT   for treating a subject infected with human acquired immune deficiency
PT   syndrome virus.
XX
PS   Disclosure; Fig 6A-B; 69pp; English.
XX
CC   The present sequence represents a human T4 glycoprotein. An aqueous-
CC   soluble polypeptide comprising a portion of a human T4 glycoprotein
CC   specifically forms a complex with a human immunodeficiency virus (HIV)
CC   envelope glycoprotein. The DNA is useful for producing the soluble
CC   surface T4 glycoprotein. The soluble surface T4 glycoprotein is useful as
CC   a therapeutic agent, i.e. as prophylaxis for treating a subject infected
CC   with an HIV virus. Thus, the soluble T4 glycoprotein is useful for
CC   treating human AIDS. The soluble T4 glycoprotein is also useful in
CC   diagnostic or screening assays, e.g. for screening inhibitors of virus
CC   binding, or for detecting and quantitating T4, T4+ cells and antibodies
CC   to T4, which are of diagnostic value for AIDS
XX
SQ   Sequence 458 AA;

  Query Match            88.0%;  Score 2046;  DB 3;  Length 458;
  Best Local Similarity  89.6%;  Pred. No. 1.7e-131;
  Matches  407;  Conservative    2;  Mismatches   15;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
```

```
Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                  ||     |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425
```

RESULT 4
AAB19507
ID AAB19507 standard; protein; 729 AA.
XX
AC AAB19507;
XX
DT 09-JAN-2001 (first entry)
XX
DE CD4-IgG1 fusion protein CH4Hgamma1.
XX
KW CD4; IgG1; human; CD4Hgamma1; fusion protein; immunoglobulin; HIV; SIV;
KW gp120; therapy; diagnosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 1. .395
FT /note= "CD4 extracellular region"
FT Protein 400. .729
FT /note= "IgG1 heavy chain"
XX
PN US6117656-A.
XX
PD 12-SEP-2000.
XX
PF 07-JUN-1995; 95US-00479353.
XX
PR 22-JAN-1988; 88US-00147351.
PR 23-JAN-1989; 89US-00299596.
PR 09-JUN-1992; 92US-00896781.
PR 12-APR-1993; 93US-00057952.
PR 04-FEB-1994; 94US-00191708.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-586558/55.
DR N-PSDB; AAA50660.
XX
PT CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
PT SIV.
XX
PS Example 1; Col 13-30; 39pp; English.
XX
CC The present sequence is that of fusion protein CD4Hgamma1 comprising the
CC extracellular portion of CD4, which binds to HIV gp120, linked at its C-
CC terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
CC encoding CD4 was linked to IgG1 DNA at the Hind3 site upstream of the CH1
CC region (see AAA50660). Fusion protein CD4Hgamma1 and a nucleic acid
CC encoding it are claimed. Also claimed are a vector comprising the nucleic
CC acid, and a method of producing the fusion protein in secreted form using
CC a transformed host cell. The fusion protein may further comprise a
CC therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
CC can be administered to an animal (including humans) for treatment of HIV
CC or SIV infection, and can also be used in assays for HIV or SIV, imaging
CC and tissue stains. IgG1 fusion proteins such as CD4Hgamma1 provide both
CC complement-mediated and cell-mediated immunity
XX
SQ Sequence 729 AA;

Query Match 87.9%; Score 2044.5; DB 3; Length 729;
Best Local Similarity 91.5%; Pred. No. 3.8e-131;
Matches 407; Conservative 3; Mismatches 22; Indels 13; Gaps 3;

```
Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PRA-----SALPAPPT------G 407
              |||||||||||||||||||||||||||||||||||  | |      |  |  |:      |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEASTKGPSVFPLAPSSKSTSGG 420

Qy        408 SALPDPQTASALPDPPAASALPAAL 432
              :|      |  |:|   |    ||
Db        421 TAALGCLVVSYFPEPVTVSWNSGAL 445
```

RESULT 5
AAP93008
ID AAP93008 standard; protein; 729 AA.
XX
AC AAP93008;
XX
DT 25-MAR-2003 (revised)
DT 02-NOV-1992 (first entry)
XX
DE Genetic construct which encodes CD4 linked to human IgG1 at the Hind3
DE site upstream of the CH1 region (fusion protein CD4H-gamma-1).
XX
KW Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
KW diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
XX
OS Homo sapiens.
XX
PN EP325262-A.
XX
PD 26-JUL-1989.
XX
PF 20-JAN-1989; 89EP-00100913.
XX
PR 22-JAN-1988; 88US-00147351.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 1989-214472/30.

```
DR   N-PSDB; AAN90356.
XX
PT   Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PT   infections or detecting HIV or SIV in sample.
XX
PS   Example; Table 1, Page 12-23; 68pp; English.
XX
CC   The fusion protein genes of the invention pref. comprise cDNA sequences
CC   which encode CD4 or a fragment which binds gp120 ligated to an expression
CC   plasmid which encodes an antibody in which the variable region of the
CC   gene has been deleted (see WO87-02671). The CD4 portion of the fusion
CC   protein may comprise the complete CD4 sequence, the 370 AA extracellular
CC   region and the membrane spanning domain, or the extracellular region. The
CC   Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are
CC   specifically claimed: fusion proteins CD4H-gamma-1,CD4Mmu,CD4Pmu,CD4E-
CC   gamma1,and CD4Mmu (No.67608),pCD4P-gamma (No.67609) and pCD4E-gamma-1
CC   (No.67610). The plasmid containing (pCD4H-gamma-1) has been deposited in
CC   E. coli (MC1061/P3) at the ATCC under accession number 67611. (Updated on
CC   25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 729 AA;

  Query Match             87.8%;  Score 2042;  DB 1;  Length 729;
  Best Local Similarity   92.3%;  Pred. No. 5.6e-131;
  Matches  405;  Conservative    3;  Mismatches    9;  Indels   22;  Gaps    4;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PRA-----SALPAPPT------G 407
              |||||||||||||||||||||||||||||||||||  | |     |  |  |:      |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEASTKGPSVFPLAPSSKSTSGG 420

Qy        408 SA---------LPDPQTAS 417
              :|          |:| | |
Db        421 TAALGCLVKDYFPEPVTVS 439


RESULT 6
AAY51078
ID   AAY51078 standard; protein; 729 AA.
XX
AC   AAY51078;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human fusion protein CD4H-1.
XX
KW   Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;
KW   anti-human immunodeficiency virus; CD4H-1.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   US6004781-A.
XX
PD   21-DEC-1999.
XX
PF   04-FEB-1994;   94US-00191708.
XX
PR   22-JAN-1988;   88US-00147351.
PR   23-JAN-1989;   89US-00299596.
PR   09-JUN-1992;   92US-00896781.
PR   12-APR-1993;   93US-00057952.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-085792/07.
DR   N-PSDB; AAZ44061.
XX
PT   Fusion protein useful for the treatment of human immunodeficiency virus.
XX
PS   Example 1; Col 15-30; 39pp; English.
XX
CC   This invention describes a novel nucleic acid (I) encoding a fusion
CC   protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
CC   and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
CC   chain (III). The products pf the invention have anti-human
CC   immunodeficiency virus (HIV) activity and are capable of binding to
CC   gp120. The fusion protein is useful for treating human immunodeficiency
CC   virus (HIV) or simian immunodeficiency virus (SIV). This sequence
CC   represents the fusion protein CD4H-1 which is constructed from CD4 linked
CC   to human IgG1 upstream of the CH1 region
XX
SQ   Sequence 729 AA;

  Query Match             87.8%;  Score 2042;  DB 3;  Length 729;
  Best Local Similarity   92.3%;  Pred. No. 5.6e-131;
  Matches  405;  Conservative    3;  Mismatches    9;  Indels   22;  Gaps    4;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PRA-----SALPAPPT------G 407
              |||||||||||||||||||||||||||||||||||  | |     |  |  |:      |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEASTKGPSVFPLAPSSKSTSGG 420

Qy        408 SA---------LPDPQTAS 417
              :|          |:| | |
Db        421 TAALGCLVKDYFPEPVTVS 439
```

```
RESULT 7
AAY59168
ID   AAY59168 standard; protein; 729 AA.
XX
AC   AAY59168;
XX
DT   14-MAR-2000  (first entry)
XX
DE   CD4-Ig fusion protein CD4Hgamma1.
XX
KW   HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
KW   secreted protein; SIV infection; medicament.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   CA1340741-C.
XX
PD   14-SEP-1999.
XX
PF   20-JAN-1989;   89CA-00588749.
XX
PR   20-JAN-1989;   89CA-00588749.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-063015/06.
DR   N-PSDB; AAZ48201.
XX
PT   New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
PT   the treatment of HIV or simian immunodeficiency virus infections.
XX
PS   Example 1; Page 25-36; 89pp; English.
XX
CC   The invention provides a fusion gene encoding a fusion protein that
CC   comprises an extracellular CD4 DNA sequence or its fragment which binds
CC   to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
CC   sequence of an Ig heavy or light chain, where the DNA sequence encoding
CC   the variable region has been replaced with the DNA sequence which encodes
CC   extracellular CD4 or its gp120 binding fragment. The fusion protein is
CC   capable of being secreted. The fusion proteins are useful for treating
CC   HIV or SIV infections in animals, preferably humans. They are also useful
CC   for producing medicaments which can be used for treating HIV or SIV
CC   infections in humans. The present sequence represents the fusion protein
CC   CD4Hgamma1 where the CD4 is linked to human IgG1 at the Hind3 site
CC   upstream of the CH1 region
XX
SQ   Sequence 729 AA;

  Query Match             87.8%;  Score 2042;  DB 3;  Length 729;
  Best Local Similarity   92.3%;  Pred. No. 5.6e-131;
  Matches  405;  Conservative    3;  Mismatches    9;  Indels   22;  Gaps    4;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PRA-----SALPAPPT------G 407
              |||||||||||||||||||||||||||||||||||  | |     |  |  |:      |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEASTKGPSVFPLAPSSKSTSGG 420

Qy        408 SA---------LPDPQTAS 417
              :|         |:| | |
Db        421 TAALGCLVKDYFPEPVTVS 439


RESULT 8
AAB19509
ID   AAB19509 standard; protein; 416 AA.
XX
AC   AAB19509;
XX
DT   09-JAN-2001  (first entry)
XX
DE   CD4-IgM fusion protein CH4Mmu.
XX
KW   CD4; IgM; human; CD4Mmu; fusion protein; immunoglobulin; HIV; SIV; gp120;
KW   therapy; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         1. .395
FT                   /note= "CD4 extracellular region"
FT   Protein         400. .416
FT                   /note= "IgM heavy chain partial sequence"
XX
PN   US6117656-A.
XX
PD   12-SEP-2000.
XX
PF   07-JUN-1995;   95US-00479353.
XX
PR   22-JAN-1988;   88US-00147351.
PR   23-JAN-1989;   89US-00299596.
PR   09-JUN-1992;   92US-00896781.
PR   12-APR-1993;   93US-00057952.
PR   04-FEB-1994;   94US-00191708.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-586558/55.
DR   N-PSDB; AAA50662.
XX
PT   CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
PT   SIV.
XX
PS   Example 1; Col 41-50; 39pp; English.
XX
CC   The present sequence is that of fusion protein CD4Mmu comprising the
CC   extracellular portion of CD4, which binds to HIV gp120, linked at its C-
CC   terminus to the human IgM heavy chain. To obtain the fusion protein, DNA
CC   encoding CD4 was linked to IgM DNA at the Mst2 site upstream of the CH1
CC   region (see AAA50662). Fusion protein CD4Mmu and a nucleic acid encoding
CC   it are claimed. Also claimed are a vector comprising the nucleic acid,
CC   and a method of producing the fusion protein in secreted form using a
CC   transformed host cell. The fusion protein may further comprise a
CC   therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
CC   can be administered to an animal (including humans) for treatment of HIV
CC   or SIV infection, and can also be used in assays for HIV or SIV, imaging
CC   and tissue stains. IgM fusion proteins such as CD4Mmu provide complement-
CC   mediated immunity
```

```
XX
SQ   Sequence 416 AA;

  Query Match             87.8%;  Score 2040.5;  DB 3;  Length 416;
  Best Local Similarity   97.6%;  Pred. No. 3.6e-131;
  Matches  399;  Conservative    0;  Mismatches    3;  Indels    7;  Gaps    1;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSA 409
              ||||||||||||||||||||||||||||||||||      | | |||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH-------ADPEGSA 402


RESULT 9
AAR27277
ID   AAR27277 standard; protein; 462 AA.
XX
AC   AAR27277;
XX
DT   25-MAR-2003  (revised)
DT   28-JUL-1995  (first entry)
XX
DE   CD4:eta peptide chimeric protein.
XX
KW   Fusion protein; CD4; extracellular domain; zeta; eta; gamma;
KW   membrane spanning domain; intracellular domain; type I;
KW   integral membrane homodimer; TCR; T cell antigen receptor;
KW   extracellular domain; mouse; human; receptor; chimera;
KW   HPB-ALL tumour cell line; natural killer cell.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         1. .399
FT                   /note= "CD4 extracellular domain"
FT   Protein         400. .462
FT                   /note= "Zeta membrane spanning and intracellular domain"
XX
PN   WO9215322-A1.
XX
PD   17-SEP-1992.
XX
PF   06-MAR-1992;   92WO-US001785.
XX
PR   07-MAR-1991;   91US-00665961.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1992-331474/40.
DR   N-PSDB; AAQ28705.
XX
PT   Therapeutic cells expressing chimeric receptors - directing cellular
PT   response to an infective agent, useful in treating HIV-1, AIDS
PT   Pneumocystis carinii infections etc.
XX
PS   Example 2; Page 73-74; 114pp; English.
XX
CC   This sequence represents a fusion protein between the CD4 extracellular
CC   domain and the eta protein membrane spanning domain and intracellular
CC   domain. Eta is an isoform of zeta (see also AAR27276) which is a 32 kD
CC   type I integral membrane homodimer, which arises by alternate mRNA
CC   splicing. It is present in reduced amounts in cells expressing the T cell
CC   antigen receptor. Zeta-eta heterodimers are thought to mediate the
CC   formation of inositol phosphates, as well as the receptor initiated cell
CC   death called apoptosis. In the production of the CD4 receptor chimera,
CC   the eta cDNA was isolated from the HPB-ALL tumour cell line and from
CC   human natural killer cells. The eta cDNA was joined to the extracellular
CC   domain of an engineered form of CD4 possessing a BamHI site just upstream
CC   of the membrane spanning domain, by a BamHI site naturally present a few
CC   residues upstream of the membrane spanning domain. (Updated on 25-MAR-
CC   2003 to correct PN field.)
XX
SQ   Sequence 462 AA;

  Query Match             87.7%;  Score 2040;  DB 2;  Length 462;
  Best Local Similarity   91.4%;  Pred. No. 4.4e-131;
  Matches  405;  Conservative    1;  Mismatches    9;  Indels   28;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||||||      |||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------------ADPQ------ 399

Qy        421 DPPAASALPAALAVISFLLGLGL 443
                    |   |  | || |: |
Db        400 -------LCYILDAILFLYGIVL 415


RESULT 10
AAR78677
ID   AAR78677 standard; protein; 462 AA.
XX
AC   AAR78677;
XX
DT   16-APR-1996  (first entry)
XX
DE   T-cell receptor gamma.
XX
```

```
KW   Chimeric receptor; CD4; T-cell receptor gamma; HIV; cytolysis;
KW   human immunodeficiency virus; adoptive immunotherapy.
XX
OS   Homo sapiens.
XX
PN   WO9521528-A1.
XX
PD   17-AUG-1995.
XX
PF   12-JAN-1995;   95WO-US000454.
XX
PR   14-FEB-1994;   94US-00195395.
PR   02-AUG-1994;   94US-00284391.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Banapour B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1995-292893/38.
DR   P-PSDB; AAQ96123.
XX
PT   Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
PT   cells.
XX
PS   Example 2; Page 77-78; 118pp; English.
XX
CC   Fusion proteins comprising the extracellular domain of CD4 fused to T-
CC   cell receptor zeta, gamma or eta (AAR78676-78, respectively) were
CC   expressed in CV1 using a vaccine virus vector. These CD4:zeta, CD4:gamma
CC   and CD4:eta chimeric receptors mediated cytolysis of targets expressing
CC   HIV gp120/41
XX
SQ   Sequence 462 AA;

  Query Match             87.7%;  Score 2040;  DB 2;  Length 462;
  Best Local Similarity   91.4%;  Pred. No. 4.4e-131;
  Matches  405;  Conservative    1;  Mismatches    9;  Indels   28;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||||||||                 |||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------------ADPQ------ 399

Qy        421 DPPAASALPAALAVISFLLGLGL 443
                      |   |  | || |: |
Db        400 -------LCYILDAILFLYGIVL 415


RESULT 11
AAR89457
```

```
ID   AAR89457 standard; protein; 462 AA.
XX
AC   AAR89457;
XX
DT   26-SEP-1996  (first entry)
XX
DE   CD4:gamma fusion protein.
XX
KW   CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
KW   human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
KW   dendritic cell; therapy; mammal; infection.
XX
OS   Synthetic.
XX
PN   WO9603883-A1.
XX
PD   15-FEB-1996.
XX
PF   26-JUL-1995;   95WO-US009468.
XX
PR   02-AUG-1994;   94US-00284391.
PR   24-FEB-1995;   95US-00394388.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Banapour B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1996-129034/13.
DR   N-PSDB; AAT10802.
XX
PT   Membrane-bound chimeric receptor comprising extracellular portion
PT   including CD4 fragment - cells expressing receptor can be used for
PT   treatment of HIV infection.
XX
PS   Example 2; Page 79; 134pp; English.
XX
CC   AAT10801-T10803 represent membrane bound proteinaceous chimeric receptors
CC   of the invention. This sequence represents the CD4:gamma chimera. The
CC   transmembrane region of the chimeric receptor acts to separate the
CC   intracellular and extracellular domains of the chimera, and contains a
CC   portion of the CD7 (see AAR89440), CD5 or CD34 transmembrane domains.
CC   Alternatively, the extracellular portion of the receptor can be separated
CC   from the intracellular domain by the hinge, CH2 and CH3 domains of human
CC   IgG1 (see AAR89441). The extracellular portion of the chimeric receptor
CC   contains a fragment of CD4 (amino acids 1-394 or 1-200 of the CD4
CC   sequence, see AAR89450 and AAR89451) which specifically recognises and
CC   binds HIV-infected cells, but does not mediate HIV infection. The
CC   extracellular domain of the receptor is separated from the cell membrane
CC   by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices. The
CC   cells expressing the receptor are preferably T cells, B cells,
CC   neutrophils, or dendritic cells. The therapeutic cells expressing the
CC   chimeric receptor are administered to a mammal to treat HIV infection
XX
SQ   Sequence 462 AA;

  Query Match             87.7%;  Score 2040;  DB 2;  Length 462;
  Best Local Similarity   91.4%;  Pred. No. 4.4e-131;
  Matches  405;  Conservative    1;  Mismatches    9;  Indels   28;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||||||||                  |||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------------ADPQ------ 399

Qy        421 DPPAASALPAALAVISFLLGLGL 443
                    |   |  | || |: |
Db        400 -------LCYILDAILFLYGIVL 415
```

RESULT 12
AAW02214
ID AAW02214 standard; protein; 462 AA.
XX
AC AAW02214;
XX
DT 11-NOV-1996 (first entry)
XX
DE CD4:Fc receptor gamma chain chimaeric receptor.
XX
KW Chimaeric receptor; cellular immunity; adoptive immunotherapy; CD4;
KW human immunodeficiency virus type 1; HIV-1; AIDS; therapy;
KW Fc receptor gamma chain; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 1. .393
FT /label= Extracellular_domain
FT /note= "CD4 extracellular domain"
FT Region 394. .397
FT /label= Linker
FT /note= "encoding DNA contains a BamHI site used for
FT fusion construction"
FT Region 398. .462
FT /note= "region of fusion derived from gamma chain,
FT preferred signal-transducing portions for constructs of
FT the invention are amino acids 421-462 and 402-419"
FT Domain 400. .462
FT /label= Transmembrane+Intracellular_domains
XX
PN WO9625953-A1.
XX
PD 29-AUG-1996.
XX
PF 25-JAN-1996; 96WO-US001056.
XX
PR 24-FEB-1995; 95US-00394176.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Seed B, Romeo C, Kolanus W;
XX
DR WPI; 1996-402134/40.
DR N-PSDB; AAT36759.
XX
PT Direction of cellular immune response using therapeutic cell expressing 2
PT chimaeric receptors - comprising region binding to target cell and region
PT that signals target cell destruction, or CD28 region, partic. for
PT eliminating HIV-infected cells.
XX
PS Claim 7; Page 76; 120pp; English.
XX
CC A chimaeric receptor (AAW00214) comprises the extracellular domain of an
CC engineered form of the CD4 cellular receptor for HIV and the
CC transmembrane and intracellular regions, including the cytolytic signal-
CC transducing portion, of the human Fc receptor gamma chain; the region of
CC the fusion is shown in AAW02223. It can be obtd. by inserting a gene
CC fusion (AAT36759) into a vaccinia virus vector and expression in host
CC cells. Chimaeric receptors comprising CD4 fused to Fc receptor gamma or T
CC -cell receptor zeta (see also AAW02213) or eta (AAW02215) chains are
CC capable of directing cytotoxic T lymphocytes to specifically recognise
CC and kill cells expressing HIV gp120, thus providing a therapy for AIDS
XX
SQ Sequence 462 AA;

```
  Query Match             87.7%;  Score 2040;  DB 2;  Length 462;
  Best Local Similarity   91.4%;  Pred. No. 4.4e-131;
  Matches  405;  Conservative    1;  Mismatches    9;  Indels   28;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||||||||                  |||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------------ADPQ------ 399

Qy        421 DPPAASALPAALAVISFLLGLGL 443
                    |   |  | || |: |
Db        400 -------LCYILDAILFLYGIVL 415
```

RESULT 13
AAW83142
ID AAW83142 standard; protein; 462 AA.
XX
AC AAW83142;
XX
DT 03-FEB-1999 (first entry)
XX
DE Chimeric receptor containing mouse gamma polypeptide.
XX
KW Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;
KW tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;
KW CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;
KW protozoan; viral.
XX
OS Synthetic.
OS Mus sp.
XX
PN US5843728-A.
XX
PD 01-DEC-1998.
XX
PF 05-APR-1995; 95US-00417495.
XX

```
PR   07-MAR-1991;   91US-00665961.
PR   06-MAR-1992;   92US-00847566.
PR   28-FEB-1994;   94US-00203866.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Romeo C,  Kolanus W,  Seed B;
XX
DR   WPI; 1999-044582/04.
DR   N-PSDB; AAV70158.
XX
PT   Membrane-bound chimeric receptors - comprising extracellular portion
PT   which recognises and binds a target cell and an intracellular portion of
PT   e.g. a T-cell receptor.
XX
PS   Example 2; Col 43-46; 57pp; English.
XX
CC   The present invention describes DNA encoding a membrane-bound chimeric
CC   receptor comprising: (a) an extracellular portion that specifically
CC   recognises and binds a target cell or a target infective agent; and (b)
CC   an intracellular portion of a T-cell receptor CD3, zeta or eta
CC   polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.
CC   The present sequence represents a chimeric receptor containing the mouse
CC   gamma polypeptide. Cells expressing chimeric receptors of the present
CC   invention can be administered to mammals in order to destroy pathogens
CC   (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells
CC   or autoimmune-generated cells
XX
SQ   Sequence 462 AA;

  Query Match             87.7%;  Score 2040;  DB 2;  Length 462;
  Best Local Similarity   91.4%;  Pred. No. 4.4e-131;
  Matches  405;  Conservative    1;  Mismatches    9;  Indels   28;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||||||||                 |||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------------ADPQ------ 399

Qy        421 DPPAASALPAALAVISFLLGLGL 443
                    |   |  | || |: |
Db        400 -------LCYILDAILFLYGIVL 415


RESULT 14
AAR27276
ID   AAR27276 standard; protein; 575 AA.
XX
AC   AAR27276;
XX
DT   25-MAR-2003  (revised)
DT   28-JUL-1995  (first entry)
XX
DE   CD4:zeta peptide chimeric protein.
XX
KW   Fusion protein; CD4; extracellular domain; zeta; eta; gamma;
KW   membrane spanning domain; intracellular domain; type I;
KW   integral membrane homodimer; TCR; T cell antigen receptor;
KW   extracellular domain; mouse; human; receptor; chimera;
KW   HPB-ALL tumour cell line; natural killer cell.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         1. .399
FT                   /note= "CD4 extracellular domain"
FT   Protein         400. .575
FT                   /note= "Zeta membrane spanning and intracellular domain"
XX
PN   WO9215322-A1.
XX
PD   17-SEP-1992.
XX
PF   06-MAR-1992;   92WO-US001785.
XX
PR   07-MAR-1991;   91US-00665961.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1992-331474/40.
DR   N-PSDB; AAQ28704.
XX
PT   Therapeutic cells expressing chimeric receptors - directing cellular
PT   response to an infective agent, useful in treating HIV-1, AIDS
PT   Pneumocystis carinii infections etc.
XX
PS   Example 2; Page 72-73; 114pp; English.
XX
CC   This sequence represents a fusion protein between the CD4 extracellular
CC   domain and the zeta protein membrane spanning domain and intracellular
CC   domain. Zeta is a 32 kD type I integral membrane homodimer which has a 9
CC   residue extracellular domain and a 112/113 residue intracellular domain
CC   for mouse and human protein respectively. In the production of the CD4
CC   receptor chimera, the zeta cDNA was isolated from the HPB-ALL tumour cell
CC   line and from human natural killer cells. The zeta cDNA was joined to the
CC   extracellular domain of an engineered form of CD4 possessing a BamHI site
CC   just upstream of the membrane spanning domain, by a BamHI site naturally
CC   present a few residues upstream of the membrane spanning domain. (Updated
CC   on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 575 AA;

  Query Match             87.7%;  Score 2038.5;  DB 2;  Length 575;
  Best Local Similarity   92.1%;  Pred. No. 7.3e-131;
  Matches  406;  Conservative    4;  Mismatches    8;  Indels   23;  Gaps    3;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
```

```
Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PR----------------ASAL 401
              ||||||||||||||||||||||||||||||||||  |:                |  |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPKLCYLLDGILFIYGVIITALYL 420

Qy        402 PAPPTGSALPDPQTASALPDP 422
              |  : ||    :||: | ||
Db        421 RAKFSRSA----ETAANLQDP 437


RESULT 15
AAR78676
ID   AAR78676 standard; protein; 575 AA.
XX
AC   AAR78676;
XX
DT   16-APR-1996  (first entry)
XX
DE   T-cell receptor zeta.
XX
KW   Chimeric receptor; CD4; T-cell receptor zeta; HIV; cytolysis;
KW   human immunodeficiency virus; adoptive immunotherapy.
XX
OS   Homo sapiens.
XX
PN   WO9521528-A1.
XX
PD   17-AUG-1995.
XX
PF   12-JAN-1995;   95WO-US000454.
XX
PR   14-FEB-1994;   94US-00195395.
PR   02-AUG-1994;   94US-00284391.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Banapour B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1995-292893/38.
DR   N-PSDB; AAQ96122.
XX
PT   Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
PT   cells.
XX
PS   Example 2; Page 76-77; 118pp; English.
XX
CC   Fusion proteins comprising the extracellular domain of CD4 fused to T-
CC   cell receptor zeta, gamma or eta (AAR78676-78, respectively) were
CC   expressed in CV1 using a vaccine virus vector. These CD4:zeta, CD4:gamma
CC   and CD4:eta chimeric receptors mediated cytolysis of targets expressing
CC   HIV gp120/41
XX
SQ   Sequence 575 AA;

  Query Match             87.7%;  Score 2038.5;  DB 2;  Length 575;
  Best Local Similarity   92.1%;  Pred. No. 7.3e-131;
  Matches  406;  Conservative    4;  Mismatches    8;  Indels   23;  Gaps    3;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PR----------------ASAL 401
              ||||||||||||||||||||||||||||||||||  |:                |  |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPKLCYLLDGILFIYGVIITALYL 420

Qy        402 PAPPTGSALPDPQTASALPDP 422
              |  : ||    :||: | ||
Db        421 RAKFSRSA----ETAANLQDP 437


RESULT 16
AAR89456
ID   AAR89456 standard; protein; 575 AA.
XX
AC   AAR89456;
XX
DT   26-SEP-1996  (first entry)
XX
DE   CD4:zeta fusion protein.
XX
KW   CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
KW   human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
KW   dendritic cell; therapy; mammal; infection.
XX
OS   Synthetic.
XX
PN   WO9603883-A1.
XX
PD   15-FEB-1996.
XX
PF   26-JUL-1995;   95WO-US009468.
XX
PR   02-AUG-1994;   94US-00284391.
PR   24-FEB-1995;   95US-00394388.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Banapour B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1996-129034/13.
DR   N-PSDB; AAT10801.
XX
PT   Membrane-bound chimeric receptor comprising extracellular portion
PT   including CD4 fragment - cells expressing receptor can be used for
PT   treatment of HIV infection.
XX
PS   Example 2; Page 77-78; 134pp; English.
XX
CC   AAT10801-T10803 represent membrane bound proteinaceous chimeric receptors
CC   of the invention. This sequence represents the CD4:zeta chimera. The
CC   transmembrane region of the chimeric receptor acts to separate the
CC   intracellular and extracellular domains of the chimera, and contains a
CC   portion of the CD7 (see AAR89440), CD5 or CD34 transmembrane domains.
CC   Alternatively, the extracellular portion of the receptor can be separated
CC   from the intracellular domain by the hinge, CH2 and CH3 domains of human
CC   IgG1 (see AAR89441). The extracellular portion of the chimeric receptor
CC   contains a fragment of CD4 (amino acids 1-394 or 1-200 of the CD4
```

```
CC   sequence, see AAR89450 and AAR89451) which specifically recognises and
CC   binds HIV-infected cells, but does not mediate HIV infection. The
CC   extracellular domain of the receptor is separated from the cell membrane
CC   by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices. The
CC   cells expressing the receptor are preferably T cells, B cells,
CC   neutrophils, or dendritic cells. The therapeutic cells expressing the
CC   chimeric receptor are administered to a mammal to treat HIV infection
XX
SQ   Sequence 575 AA;

  Query Match             87.7%;  Score 2038.5;  DB 2;  Length 575;
  Best Local Similarity   92.1%;  Pred. No. 7.3e-131;
  Matches  406;  Conservative    4;  Mismatches    8;  Indels   23;  Gaps    3;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PR-----------------ASAL 401
              ||||||||||||||||||||||||||||||||||  |:                 | |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPKLCYLLDGILFIYGVIITALYL 420

Qy        402 PAPPTGSALPDPQTASALPDP 422
              |  : ||    :||: | ||
Db        421 RAKFSRSA----ETAANLQDP 437


RESULT 17
AAW02213
ID   AAW02213 standard; protein; 575 AA.
XX
AC   AAW02213;
XX
DT   11-NOV-1996  (first entry)
XX
DE   CD4:T-cell receptor zeta chain chimaeric receptor.
XX
KW   Chimaeric receptor; cellular immunity; adoptive immunotherapy; CD4;
KW   human immunodeficiency virus type 1; HIV-1; AIDS; therapy;
KW   T-cell receptor zeta chain; cytotoxic T lymphocyte; CTL.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1. .393
FT                   /label= Extracellular_domain
FT                   /note= "CD4 extracellular domain"
FT   Region          394. .396
FT                   /label= Linker
FT                   /note= "encoding DNA contains a BamHI site used for
FT                   fusion construction"
FT   Region          397. .575
FT                   /note= "region of fusion derived from zeta chain,
FT                   preferred signal-transducing portions for constructs of
FT                   the invention are amino acids 421-575, 423-255, 438-455,
FT                   461-494, 494-528, 400-420 and 421-462"
FT   Domain          400. .437
FT                   /label= Transmembrane_domain
FT                   /note= "zeta chain transmembrane domain"
FT   Domain          438. .575
FT                   /label= Intracellular_domain
FT                   /note= "zeta chain intracellular domain"
XX
PN   WO9625953-A1.
XX
PD   29-AUG-1996.
XX
PF   25-JAN-1996;   96WO-US001056.
XX
PR   24-FEB-1995;   95US-00394176.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1996-402134/40.
DR   N-PSDB; AAT36758.
XX
PT   Direction of cellular immune response using therapeutic cell expressing 2
PT   chimaeric receptors - comprising region binding to target cell and region
PT   that signals target cell destruction, or CD28 region, partic. for
PT   eliminating HIV-infected cells.
XX
PS   Claim 7; Page 74-75; 120pp; English.
XX
CC   A chimaeric receptor (AAW00213) comprises the extracellular domain of an
CC   engineered form of the CD4 cellular receptor for HIV and the
CC   transmembrane and intracellular regions, including the cytolytic signal-
CC   transducing portion, of the human T-cell receptor zeta chain; the region
CC   of the fusion is shown in AAW02221. It can be obtd. by inserting a gene
CC   fusion (AAT36758) into a vaccinia virus vector and expression in host
CC   cells. Chimaeric receptors comprising CD4 fused to zeta, eta (see also
CC   AAW02215) or Fc receptor gamma (see also AAW02214) chains are capable of
CC   directing cytotoxic T lymphocytes to specifically recognise and kill
CC   cells expressing HIV gp120, thus providing a therapy for AIDS
XX
SQ   Sequence 575 AA;

  Query Match             87.7%;  Score 2038.5;  DB 2;  Length 575;
  Best Local Similarity   92.1%;  Pred. No. 7.3e-131;
  Matches  406;  Conservative    4;  Mismatches    8;  Indels   23;  Gaps    3;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
```

```
Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PR----------------ASAL 401
              ||||||||||||||||||||||||||||||||||||  |:                |  |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPKLCYLLDGILFIYGVIITALYL 420

Qy        402 PAPPTGSALPDPQTASALPDP 422
               |  : ||    :||: | ||
Db        421 RAKFSRSA----ETAANLQDP 437


RESULT 18
AAW83140
ID   AAW83140 standard; protein; 575 AA.
XX
AC   AAW83140;
XX
DT   03-FEB-1999  (first entry)
XX
DE   Chimeric receptor containing human zeta polypeptide.
XX
KW   Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;
KW   tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;
KW   CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;
KW   protozoan; viral.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   US5843728-A.
XX
PD   01-DEC-1998.
XX
PF   05-APR-1995;   95US-00417495.
XX
PR   07-MAR-1991;   91US-00665961.
PR   06-MAR-1992;   92US-00847566.
PR   28-FEB-1994;   94US-00203866.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Romeo C,  Kolanus W,  Seed B;
XX
DR   WPI; 1999-044582/04.
DR   N-PSDB; AAV70156.
XX
PT   Membrane-bound chimeric receptors - comprising extracellular portion
PT   which recognises and binds a target cell and an intracellular portion of
PT   e.g. a T-cell receptor.
XX
PS   Example 2; Col 39-42; 57pp; English.
XX
CC   The present invention describes DNA encoding a membrane-bound chimeric
CC   receptor comprising: (a) an extracellular portion that specifically
CC   recognises and binds a target cell or a target infective agent; and (b)
CC   an intracellular portion of a T-cell receptor CD3, zeta or eta
CC   polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.
CC   The present sequence represents a chimeric receptor containing the human
CC   zeta polypeptide. Cells expressing chimeric receptors of the present
CC   invention can be administered to mammals in order to destroy pathogens
CC   (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells
CC   or autoimmune-generated cells
XX
SQ   Sequence 575 AA;

  Query Match             87.7%;  Score 2038.5;  DB 2;  Length 575;
  Best Local Similarity   92.1%;  Pred. No. 7.3e-131;
  Matches  406;  Conservative    4;  Mismatches    8;  Indels   23;  Gaps    3;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PR----------------ASAL 401
              ||||||||||||||||||||||||||||||||||||  |:                |  |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPKLCYLLDGILFIYGVIITALYL 420

Qy        402 PAPPTGSALPDPQTASALPDP 422
               |  : ||    :||: | ||
Db        421 RAKFSRSA----ETAANLQDP 437


RESULT 19
AAP81990
ID   AAP81990 standard; protein; 458 AA.
XX
AC   AAP81990;
XX
DT   25-MAR-2003  (revised)
DT   15-OCT-1990  (first entry)
XX
DE   Clone pT4B encoded HIV T4 glycoprotein.
XX
KW   Human Immunodeficiency Virus T4 envelope glycoprotein; AIDS; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= leader peptide
FT   Region          24. .117
FT                   /label= variable-like region 1
FT   Region          118. .134
FT                   /label= joining-like region 1
FT   Region          135. .189
FT                   /label= variable-like region 2
FT   Region          190. .205
FT                   /label= joining-like region 2
FT   Region          206. .287
FT                   /label= variable-like region 3
FT   Region          288. .309
FT                   /label= joining-like region 3
FT   Modified-site   297. .299
FT                   /label= putative N-glycosylation site
FT   Region          310. .377
FT                   /label= variable-like region 4
FT   Modified-site   325. .327
FT                   /label= putative N-glycosylation site
FT   Region          378. .397
FT                   /label= joining-like region 4
FT   Region          398. .420
FT                   /label= trans-membrane region
FT   Region          421. .458
FT                   /label= cytoplasmic region
XX
PN   WO8801304-A.
XX
PD   25-FEB-1988.
```

```
XX
PF   20-AUG-1987;   87WO-US002050.
XX
PR   21-AUG-1986;   86US-00898587.
XX
PA   (UYCO-) COLUMBIA UNIV.
PA   (MADD/) MADDON P J.
XX
PI   Littman DR,  Maddon PJ,  Chess L,  Axel R,  Weiss R,  Mcdougal JS;
XX
DR   WPI; 1988-064019/09.
DR   N-PSDB; AAN80512.
XX
PT   Nucleic acid encoding T4 glyco:protein - used for treatment of AIDS and
PT   producing antibodies for use as vaccine for immunisation against AIDS.
XX
PS   Disclosure; Page ?; 128pp; English.
XX
CC   T4 protein encoded by part of 3kb insert from human T cell library
CC   (pT4B). (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 458 AA;

  Query Match           87.7%;  Score 2038;  DB 1;  Length 458;
  Best Local Similarity   89.4%;  Pred. No. 6e-131;
  Matches  406;  Conservative    2;  Mismatches   16;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||     |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 20
AAP91369
ID   AAP91369 standard; protein; 458 AA.
XX
AC   AAP91369;
XX
DT   24-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   08-JAN-1990  (first entry)
XX
DE   T4 protein.
XX
KW   T4 protein; human immunodeficiency virus; AIDS; clone pT4B.
XX
OS   Homo sapiens; (human).
XX
FH   Key             Location/Qualifiers
FT   Domain          1. .23
FT   Domain          24. .117
FT   Region          41
FT   Region          109
FT   Domain          118. .134
FT   Domain          135. .189
FT   Region          155
FT   Region          184
FT   Domain          190. .204
FT   Domain          205. .286
FT   Domain          287. .309
FT   Binding-site    296. .298
FT   Domain          310. .376
FT   Binding-site    325. .327
FT   Region          328
FT   Region          370
FT   Domain          377. .397
FT   Domain          398. .420
FT   Domain          421. .458
XX
PN   EP330227-A.
XX
PD   30-AUG-1989.
XX
PF   24-FEB-1989;   89EP-00103297.
XX
PR   24-FEB-1988;   88US-00160348.
XX
PA   (UYCO-) COLUMBIA UNIV.
PA   (SMIK ) SMITHKLINE BECKMAN CORP.
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Maddon PJ,  Axel R,  Sweet RW,  Arthos J;
XX
DR   WPI; 1989-250337/35.
XX
PT   Soluble T4 polypeptide derivs. - inhibitors of extracellular and cell to
PT   cell spread of HIV used in prevention and treatment of AIDS.
XX
PS   Claim 1; Fig 6; 73pp; English.
XX
CC   T4 protein (AAN90619) inhibits extracellular and cell-to-cell spread of
CC   HIV. The therapeutic agent consists of amino acids +3-+185 fused to +351-
CC   +369; +3-+106 fused to +351-+369; or +3-+185. Also used to identify
CC   inhibitors of T4+ interactions, as target carrier proteins, and to
CC   generate monoclonal antibodies. Above features are: Domain 1 (D; starting
CC   at the N-terminal) = leader; D2 = variable-like-1; D3 = joining-like-1;
CC   D4 = V2; D5 = J2; D5 = V3; D6 = J3; D7 = V4; D8 = J4; D9 = transmembrane;
CC   D10 = cytoplasmic; Regions are extracellular cysteines; and the 2 sites
CC   are potential N-linked glycosylation sites. (Updated on 25-MAR-2003 to
CC   correct PA field.) (Updated on 24-OCT-2003 to standardise OS field)
XX
SQ   Sequence 458 AA;

  Query Match           87.7%;  Score 2038;  DB 1;  Length 458;
  Best Local Similarity   89.4%;  Pred. No. 6e-131;
  Matches  406;  Conservative    2;  Mismatches   16;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP----------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||     |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 21
AAY39826
ID   AAY39826 standard; protein; 458 AA.
XX
AC   AAY39826;
XX
DT   03-DEC-1999  (first entry)
XX
DE   Soluble human T4 protein.
XX
KW   Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;
KW   vaccine; immunisation; therapy.
XX
OS   Homo sapiens.
XX
PN   US5958678-A.
XX
PD   28-SEP-1999.
XX
PF   12-DEC-1994;   94US-00354452.
XX
PR   21-AUG-1986;   86US-00898587.
PR   11-JUN-1991;   91US-00713564.
PR   06-JUL-1992;   92US-00909021.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Mcdougal JS,  Weiss R,  Axel R,  Littman DR,  Maddon PJ,  Chess L;
XX
DR   WPI; 1999-561025/47.
DR   N-PSDB; AAZ20695.
XX
PT   Human T4 protein inhibits HIV binding to T4 cells, useful for treating
PT   AIDS.
XX
PS   Example 3; Fig 6; 58pp; English.
XX
CC   This sequence represents the soluble human T4 protein of the invention.
CC   The soluble human T4 protein blocks the binding of HIV to T4+ cells and
CC   is therefore useful for the treatment of AIDS. Monoclonal antibodies
CC   against the T4 protein may be used as vaccines for immunising subjects
CC   against AIDS
XX
SQ   Sequence 458 AA;

  Query Match             87.7%;  Score 2038;  DB 2;  Length 458;
  Best Local Similarity   89.4%;  Pred. No. 6e-131;
  Matches  406;  Conservative    2;  Mismatches   16;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP----------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||     |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 22
AAB19510
ID   AAB19510 standard; protein; 481 AA.
XX
AC   AAB19510;
XX
DT   09-JAN-2001  (first entry)
XX
DE   CD4-IgM fusion protein CH4Pmu.
XX
KW   CD4; IgM; human; CD4Pmu; fusion protein; immunoglobulin; HIV; SIV; gp120;
KW   therapy; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         1. .395
FT                   /note= "CD4 extracellular region"
FT   Protein         400. .481
FT                   /note= "IgM heavy chain partial sequence"
XX
PN   US6117656-A.
XX
PD   12-SEP-2000.
XX
PF   07-JUN-1995;   95US-00479353.
XX
PR   22-JAN-1988;   88US-00147351.
PR   23-JAN-1989;   89US-00299596.
PR   09-JUN-1992;   92US-00896781.
PR   12-APR-1993;   93US-00057952.
PR   04-FEB-1994;   94US-00191708.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-586558/55.
DR   N-PSDB; AAA50662.
XX
PT   CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
```

```
PT   SIV.
XX
PS   Example 1; Col 49-60; 39pp; English.
XX
CC   The present sequence is that of fusion protein CD4Pmu comprising the
CC   extracellular portion of CD4, which binds to HIV gp120, linked at its C-
CC   terminus to the human IgM heavy chain. To obtain the fusion protein, DNA
CC   encoding CD4 was linked to IgM DNA at the Pst site upstream of the CH2
CC   region (see AAA50663). Fusion protein CD4Pmu and a nucleic acid encoding
CC   it are claimed. Also claimed are a vector comprising the nucleic acid,
CC   and a method of producing the fusion protein in secreted form using a
CC   transformed host cell. The fusion protein may further comprise a
CC   therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
CC   can be administered to an animal (including humans) for treatment of HIV
CC   or SIV infection, and can also be used in assays for HIV or SIV, imaging
CC   and tissue stains. IgM fusion proteins such as CD4Pmu provide complement-
CC   mediated immunity
XX
SQ   Sequence 481 AA;

  Query Match             87.7%;  Score 2038;  DB 3;  Length 481;
  Best Local Similarity   93.2%;  Pred. No. 6.4e-131;
  Matches  400;  Conservative    4;  Mismatches    9;  Indels   16;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--------------ADPEVIAELP 405

Qy        421 DPPAASALP 429
                |  :  :|
Db        406 -PKVSVFVP 413


RESULT 23
AAY59171
ID   AAY59171 standard; protein; 481 AA.
XX
AC   AAY59171;
XX
DT   14-MAR-2000  (first entry)
XX
DE   CD4-Ig fusion protein CD4Pmu.
XX
KW   HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
KW   secreted protein; SIV infection; medicament.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   CA1340741-C.
XX
PD   14-SEP-1999.
XX
PF   20-JAN-1989;   89CA-00588749.
XX
PR   20-JAN-1989;   89CA-00588749.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-063015/06.
DR   N-PSDB; AAZ48204.
XX
PT   New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
PT   the treatment of HIV or simian immunodeficiency virus infections.
XX
PS   Example 1; Page 54-60; 89pp; English.
XX
CC   The invention provides a fusion gene encoding a fusion protein that
CC   comprises an extracellular CD4 DNA sequence or its fragment which binds
CC   to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
CC   sequence of an Ig heavy or light chain, where the DNA sequence encoding
CC   the variable region has been replaced with the DNA sequence which encodes
CC   extracellular CD4 or its gp120 binding fragment. The fusion protein is
CC   capable of being secreted. The fusion proteins are useful for treating
CC   HIV or SIV infections in animals, preferably humans. They are also useful
CC   for producing medicaments which can be used for treating HIV or SIV
CC   infections in humans. The present sequence represents the fusion protein
CC   CD4Pmu where the CD4 is linked to human IgG1 at the Pst site upstream of
CC   the CH2 region
XX
SQ   Sequence 481 AA;

  Query Match             87.7%;  Score 2038;  DB 3;  Length 481;
  Best Local Similarity   93.2%;  Pred. No. 6.4e-131;
  Matches  400;  Conservative    4;  Mismatches    9;  Indels   16;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--------------ADPEVIAELP 405

Qy        421 DPPAASALP 429
                |  :  :|
Db        406 -PKVSVFVP 413


RESULT 24
```

```
AAR89450
ID   AAR89450 standard; peptide; 398 AA.
XX
AC   AAR89450;
XX
DT   26-SEP-1996  (first entry)
XX
DE   CD4 D1-D4 domains.
XX
KW   CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
KW   human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
KW   dendritic cell; therapy; mammal; infection.
XX
OS   Homo sapiens.
XX
PN   WO9603883-A1.
XX
PD   15-FEB-1996.
XX
PF   26-JUL-1995;   95WO-US009468.
XX
PR   02-AUG-1994;   94US-00284391.
PR   24-FEB-1995;   95US-00394388.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Banapour B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1996-129034/13.
DR   N-PSDB; AAT10797.
XX
PT   Membrane-bound chimeric receptor comprising extracellular portion
PT   including CD4 fragment - cells expressing receptor can be used for
PT   treatment of HIV infection.
XX
PS   Example 10; Fig 23; 134pp; English.
XX
CC   This sequence represents the D1-D4 domains of CD4. This sequence is
CC   included in the membrane bound proteinaceous chimeric receptor of the
CC   invention. The extracellular portion of the chimeric receptor contains a
CC   fragment of CD4 (amino acids 1-394 or 1-200 of the CD4 sequence) which
CC   specifically recognises and binds HIV-infected cells, but does not
CC   mediate HIV infection. The extracellular domain of the receptor is
CC   separated from the cell membrane by 48 or 72 angstroms, or by one or more
CC   proteinaceous alpha-helices. The transmembrane region of the chimeric
CC   receptor contains a portion of the CD7, CD5 or CD34 transmembrane domain.
CC   Alternatively, the extracellular portion of the receptor can also be
CC   separated from the intracellular domain by the hinge, CH2 and CH3 domains
CC   of human IgG1. The cells expressing the receptor are preferably T cells,
CC   B cells, neutrophils, or dendritic cells. The therapeutic cells
CC   expressing the chimeric receptor are administered to a mammal to treat
CC   HIV infection
XX
SQ   Sequence 398 AA;

  Query Match             87.6%;  Score 2037;  DB 2;  Length 398;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-131;
  Matches  395;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 25
AAR27278
ID   AAR27278 standard; protein; 532 AA.
XX
AC   AAR27278;
XX
DT   25-MAR-2003  (revised)
DT   28-JUL-1995  (first entry)
XX
DE   CD4:gamma peptide chimeric protein.
XX
KW   Fusion protein; CD4; extracellular domain; zeta; eta; gamma;
KW   membrane spanning domain; intracellular domain; type I;
KW   integral membrane homodimer; TCR; T cell antigen receptor;
KW   extracellular domain; mouse; human; receptor; chimera;
KW   HPB-ALL tumour cell line; natural killer cell.
XX
OS   Homo sapiens.
XX
PN   WO9215322-A1.
XX
PD   17-SEP-1992.
XX
PF   06-MAR-1992;   92WO-US001785.
XX
PR   07-MAR-1991;   91US-00665961.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1992-331474/40.
DR   N-PSDB; AAQ28706.
XX
PT   Therapeutic cells expressing chimeric receptors - directing cellular
PT   response to an infective agent, useful in treating HIV-1, AIDS
PT   Pneumocystis carinii infections etc.
XX
PS   Example 2; Page 74-76; 114pp; English.
XX
CC   This sequence represents a fusion protein between the CD4 extracellular
CC   domain and the gamma protein membrane spanning domain and intracellular
CC   domain. The Fc-receptor-associated gamma chain is expressed in cell
CC   surface complexes with additional polypeptides, some of which mediate
CC   ligand recognition, and others which have undefined function. Gamma bears
CC   a homodimeric structure and overall organisation very similar to that of
CC   zeta (see also AAQ28704), and is a component of both the mast
CC   cell/basophil high affinity IgE receptor, Fc-epsilon-RI, which consists
CC   of at least three distinct polypeptide chains and one of the low affinity
CC   receptors for IgG, represented in mice by Fc-gamma-RII-alpha. In the
CC   production of the CD4 receptor chimera, the gamma cDNA was isolated from
CC   the HPB-ALL tumour cell line and from human natural killer cells. The
CC   gamma cDNA was joined to the extracellular domain by engineering a BamHI
CC   site just upstream of the membrane spanning domain, by a BamHI site
CC   naturally present a few residues upstream of the membrane spanning
CC   domain. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 532 AA;
```

```
  Query Match             87.6%;  Score 2037;  DB 2;  Length 532;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-131;
  Matches  395;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 26
AAR78678
ID   AAR78678 standard; protein; 532 AA.
XX
AC   AAR78678;
XX
DT   16-APR-1996  (first entry)
XX
DE   T-cell receptor eta.
XX
KW   Chimeric receptor; CD4; T-cell receptor eta; HIV; cytolysis;
KW   human immunodeficiency virus; adoptive immunotherapy.
XX
OS   Homo sapiens.
XX
PN   WO9521528-A1.
XX
PD   17-AUG-1995.
XX
PF   12-JAN-1995;   95WO-US000454.
XX
PR   14-FEB-1994;   94US-00195395.
PR   02-AUG-1994;   94US-00284391.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Banapour B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1995-292893/38.
DR   N-PSDB; AAQ96124.
XX
PT   Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
PT   cells.
XX
PS   Example 2; Page 78-79; 118pp; English.
XX
CC   Fusion proteins comprising the extracellular domain of CD4 fused to T-
CC   cell receptor zeta, gamma or eta (AAR78676-78, respectively) were
CC   expressed in CV1 using a vaccine virus vector. These CD4:zeta, CD4:gamma
CC   and CD4:eta chimeric receptors mediated cytolysis of targets expressing
CC   HIV gp120/41
XX
SQ   Sequence 532 AA;

  Query Match             87.6%;  Score 2037;  DB 2;  Length 532;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-131;
  Matches  395;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 27
AAR89458
ID   AAR89458 standard; protein; 532 AA.
XX
AC   AAR89458;
XX
DT   26-SEP-1996  (first entry)
XX
DE   CD4:eta fusion protein.
XX
KW   CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
KW   human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
KW   dendritic cell; therapy; mammal; infection.
XX
OS   Synthetic.
XX
PN   WO9603883-A1.
XX
PD   15-FEB-1996.
XX
PF   26-JUL-1995;   95WO-US009468.
XX
PR   02-AUG-1994;   94US-00284391.
PR   24-FEB-1995;   95US-00394388.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Banapour B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1996-129034/13.
DR   N-PSDB; AAT10803.
XX
PT   Membrane-bound chimeric receptor comprising extracellular portion
PT   including CD4 fragment - cells expressing receptor can be used for
PT   treatment of HIV infection.
```

```
XX
PS   Example 2; Page 80-81; 134pp; English.
XX
CC   AAT10801-T10803 represent membrane bound proteinaceous chimeric receptors
CC   of the invention. This sequence represents the CD4:eta chimera. The
CC   transmembrane region of the chimeric receptor acts to separate the
CC   intracellular and extracellular domains of the chimera, and contains a
CC   portion of the CD7 (see AAR89440), CD5 or CD34 transmembrane domains.
CC   Alternatively, the extracellular portion of the receptor can be separated
CC   from the intracellular domain by the hinge, CH2 and CH3 domains of human
CC   IgG1 (see AAR89441). The extracellular portion of the chimeric receptor
CC   contains a fragment of CD4 (amino acids 1-394 or 1-200 of the CD4
CC   sequence, see AAR89450 and AAR89451) which specifically recognises and
CC   binds HIV-infected cells, but does not mediate HIV infection. The
CC   extracellular domain of the receptor is separated from the cell membrane
CC   by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices. The
CC   cells expressing the receptor are preferably T cells, B cells,
CC   neutrophils, or dendritic cells. The therapeutic cells expressing the
CC   chimeric receptor are administered to a mammal to treat HIV infection
XX
SQ   Sequence 532 AA;

  Query Match             87.6%;  Score 2037;  DB 2;  Length 532;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-131;
  Matches  395;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 28
AAW02215
ID   AAW02215 standard; protein; 532 AA.
XX
AC   AAW02215;
XX
DT   16-OCT-2003  (revised)
DT   11-NOV-1996  (first entry)
XX
DE   CD4:T-cell receptor eta chain chimaeric receptor.
XX
KW   Chimaeric receptor; cellular immunity; adoptive immunotherapy; CD4;
KW   human immunodeficiency virus type 1; HIV-1; AIDS; therapy;
KW   T-cell receptor eta chain; cytotoxic T lymphocyte; CTL.
XX
OS   Homo; sapiens.
OS   Mus sp.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   Domain          1. .393
FT                   /label= Extracellular_domain
FT                   /note= "CD4 extracellular domain"
FT   Region          394. .396
FT                   /label= Linker
FT                   /note= "encoding DNA contains a BamHI site used for
FT                   fusion construction"
FT   Region          397. .532
FT                   /note= "region of fusion derived from eta chain,
FT                   preferred signal-transducing portions for constructs of
FT                   the invention are amino acids 421-532, 423-455, 438-455,
FT                   461-494, 494-528 or 400-420"
FT   Domain          400. .437
FT                   /label= Transmembrane_domain
FT                   /note= "eta chain transmembrane domain"
FT   Domain          438. .575
FT                   /label= Intracellular_domain
FT                   /note= "eta chain intracellular domain"
XX
PN   WO9625953-A1.
XX
PD   29-AUG-1996.
XX
PF   25-JAN-1996;   96WO-US001056.
XX
PR   24-FEB-1995;   95US-00394176.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1996-402134/40.
DR   N-PSDB; AAT36760.
XX
PT   Direction of cellular immune response using therapeutic cell expressing 2
PT   chimaeric receptors - comprising region binding to target cell and region
PT   that signals target cell destruction, or CD28 region, partic. for
PT   eliminating HIV-infected cells.
XX
PS   Claim 7; Page 77-78; 120pp; English.
XX
CC   A chimaeric receptor (AAW00215) comprises the extracellular domain of an
CC   engineered form of the CD4 cellular receptor for HIV and the
CC   transmembrane and intracellular regions, including the cytolytic signal-
CC   transducing portion, of the mouse T-cell receptor eta chain. It can be
CC   obtd. by inserting a gene fusion (AAT36760) into a vaccinia virus vector
CC   and expression in host cells. Chimaeric receptors comprising CD4 fused to
CC   eta, eta (see also AAW02213) or Fc receptor gamma (see also AAW02214)
CC   chains are capable of directing cytotoxic T lymphocytes to specifically
CC   recognise and kill cells expressing HIV gp120, thus providing a therapy
CC   for AIDS. (Updated on 16-OCT-2003 to standardise OS field)
XX
SQ   Sequence 532 AA;

  Query Match             87.6%;  Score 2037;  DB 2;  Length 532;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-131;
  Matches  395;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
                ||||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 29
AAW83141
ID   AAW83141 standard; protein; 532 AA.
XX
AC   AAW83141;
XX
DT   03-FEB-1999  (first entry)
XX
DE   Chimeric receptor containing human eta polypeptide.
XX
KW   Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;
KW   tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;
KW   CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;
KW   protozoan; viral.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   US5843728-A.
XX
PD   01-DEC-1998.
XX
PF   05-APR-1995;   95US-00417495.
XX
PR   07-MAR-1991;   91US-00665961.
PR   06-MAR-1992;   92US-00847566.
PR   28-FEB-1994;   94US-00203866.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Romeo C,  Kolanus W,  Seed B;
XX
DR   WPI; 1999-044582/04.
DR   N-PSDB; AAV70157.
XX
PT   Membrane-bound chimeric receptors - comprising extracellular portion
PT   which recognises and binds a target cell and an intracellular portion of
PT   e.g. a T-cell receptor.
XX
PS   Claim 11; Col 45-48; 57pp; English.
XX
CC   The present invention describes DNA encoding a membrane-bound chimeric
CC   receptor comprising: (a) an extracellular portion that specifically
CC   recognises and binds a target cell or a target infective agent; and (b)
CC   an intracellular portion of a T-cell receptor CD3, zeta or eta
CC   polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.
CC   The present sequence represents a chimeric receptor containing the human
CC   eta polypeptide. Cells expressing chimeric receptors of the present
CC   invention can be administered to mammals in order to destroy pathogens
CC   (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells
CC   or autoimmune-generated cells
XX
SQ   Sequence 532 AA;

  Query Match             87.6%;  Score 2037;  DB 2;  Length 532;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-131;
  Matches  395;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
                ||||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 30
AAY51082
ID   AAY51082 standard; protein; 616 AA.
XX
AC   AAY51082;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human fusion protein CD4Bgamma1.
XX
KW   Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;
KW   anti-human immunodeficiency virus; CD4Bgamma1.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   US6004781-A.
XX
PD   21-DEC-1999.
XX
PF   04-FEB-1994;   94US-00191708.
XX
PR   22-JAN-1988;   88US-00147351.
PR   23-JAN-1989;   89US-00299596.
PR   09-JUN-1992;   92US-00896781.
PR   12-APR-1993;   93US-00057952.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-085792/07.
DR   N-PSDB; AAZ44065.
XX
PT   Fusion protein useful for the treatment of human immunodeficiency virus.
XX
PS   Example 1; Col 59-70; 39pp; English.
XX
CC   This invention describes a novel nucleic acid (I) encoding a fusion
CC   protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
CC   and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
CC   chain (III). The products pf the invention have anti-human
CC   immunodeficiency virus (HIV) activity and are capable of binding to
```

```
CC   gp120. The fusion protein is useful for treating human immunodeficiency
CC   virus (HIV) or simian immunodeficiency virus (SIV). This sequence
CC   represents the fusion protein CD4Egamma1 which is constructed from CD4
CC   linked to human IgG1 upstream of the hinge region
XX
SQ   Sequence 616 AA;

  Query Match             87.6%;  Score 2037;  DB 3;  Length 616;
  Best Local Similarity   96.6%;  Pred. No. 1e-130;
  Matches  399;  Conservative    1;  Mismatches   11;  Indels    2;  Gaps    1;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PRASALPAPPTGSALP 411
              ||||||||||||||||||||||||||||||||  | |  |   |:    |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEAPELLGGPSVFLFP 413


RESULT 31
AAY59172
ID   AAY59172 standard; protein; 616 AA.
XX
AC   AAY59172;
XX
DT   14-MAR-2000  (first entry)
XX
DE   CD4-Ig fusion protein CD4Bgamma1.
XX
KW   HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
KW   secreted protein; SIV infection; medicament.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   CA1340741-C.
XX
PD   14-SEP-1999.
XX
PF   20-JAN-1989;   89CA-00588749.
XX
PR   20-JAN-1989;   89CA-00588749.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-063015/06.
DR   N-PSDB; AAZ48205.
XX
PT   New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
PT   the treatment of HIV or simian immunodeficiency virus infections.
XX
PS   Example 1; Page 61-68; 89pp; English.
XX
CC   The invention provides a fusion gene encoding a fusion protein that
CC   comprises an extracellular CD4 DNA sequence or its fragment which binds
CC   to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
CC   sequence of an Ig heavy or light chain, where the DNA sequence encoding
CC   the variable region has been replaced with the DNA sequence which encodes
CC   extracellular CD4 or its gp120 binding fragment. The fusion protein is
CC   capable of being secreted. The fusion proteins are useful for treating
CC   HIV or SIV infections in animals, preferably humans. They are also useful
CC   for producing medicaments which can be used for treating HIV or SIV
CC   infections in humans. The present sequence represents the fusion protein
CC   CD4Bgamma1 where the CD4 is linked to human IgG1 at the Ban1 site
CC   downstream from the hinge region
XX
SQ   Sequence 616 AA;

  Query Match             87.6%;  Score 2037;  DB 3;  Length 616;
  Best Local Similarity   96.6%;  Pred. No. 1e-130;
  Matches  399;  Conservative    1;  Mismatches   11;  Indels    2;  Gaps    1;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PRASALPAPPTGSALP 411
              ||||||||||||||||||||||||||||||||  | |  |   |:    |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEAPELLGGPSVFLFP 413


RESULT 32
AAP93009
ID   AAP93009 standard; protein; 631 AA.
XX
AC   AAP93009;
XX
DT   25-MAR-2003  (revised)
DT   02-NOV-1992  (first entry)
XX
DE   Genetic construct which encodes CD4 linked to human IgG1 at the Esp site
DE   upstream of the hinge region (fusion protein CD4E-gamma-1).
XX
KW   Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
KW   diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
XX
OS   Homo sapiens.
XX
PN   EP325262-A.
XX
PD   26-JUL-1989.
XX
PF   20-JAN-1989;   89EP-00100913.
```

```
XX
PR   22-JAN-1988;   88US-00147351.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 1989-214472/30.
DR   N-PSDB; AAN90357.
XX
PT   Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PT   infections or detecting HIV or SIV in sample.
XX
PS   Example; Table 2, Page 24-33; 68pp; English.
XX
CC   The fusion protein genes of the invention pref. comprise cDNA sequences
CC   which encode CD4 or a fragment which binds gp120 ligated to an expression
CC   plasmid which encodes an antibody in which the variable region of the
CC   gene has been deleted (see WO87-02671). The CD4 portion of the fusion
CC   protein may comprise the complete CD4 sequence, the 370 AA extracellular
CC   region and the membrane spanning domain, or the extracellular region. The
CC   Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are
CC   specifically claimed: fusion proteins CD4H-gamma-1,CD4Mmu,CD4Pmu,CD4E-
CC   gamma1,and CD4Mmu (No.67608),pCD4P-gamma (No.67609) and pCD4E-gamma-1
CC   (No.67610). The plasmid containing (pCD4E-gamma-1) has been deposited in
CC   E. coli (MC1061/P3) at the ATCC under accession number 67610. (Updated on
CC   25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 631 AA;

  Query Match             87.5%;  Score 2034.5;  DB 1;  Length 631;
  Best Local Similarity   92.8%;  Pred. No. 1.5e-130;
  Matches  400;  Conservative    1;  Mismatches   21;  Indels    9;  Gaps    1;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------PRASALPAPPTGSALP 411
              ||||||||||||||||||||||||||||||||||          :    |  |    |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEEPKSCDKTHTCPPCPAPELLG 420

Qy        412 DPQTASALPDP 422
              |  |   | |
Db        421 GPSVFLFPPKP 431


RESULT 33
AAB19508
ID   AAB19508 standard; protein; 631 AA.
XX
AC   AAB19508;
XX
DT   09-JAN-2001  (first entry)
XX
DE   CD4-IgG1 fusion protein CH4Egamma1.
XX
KW   CD4; IgG1; human; CD4Egamma1; fusion protein; immunoglobulin; HIV; SIV;
KW   gp120; therapy; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         1. .395
FT                   /note= "CD4 extracellular region"
FT   Protein         400. .631
FT                   /note= "IgG1 heavy chain"
XX
PN   US6117656-A.
XX
PD   12-SEP-2000.
XX
PF   07-JUN-1995;   95US-00479353.
XX
PR   22-JAN-1988;   88US-00147351.
PR   23-JAN-1989;   89US-00299596.
PR   09-JUN-1992;   92US-00896781.
PR   12-APR-1993;   93US-00057952.
PR   04-FEB-1994;   94US-00191708.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-586558/55.
DR   N-PSDB; AAA50661.
XX
PT   CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
PT   SIV.
XX
PS   Example 1; Col 29-42; 39pp; English.
XX
CC   The present sequence is that of fusion protein CD4Egamma1 comprising the
CC   extracellular portion of CD4, which binds to HIV gp120, linked at its C-
CC   terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
CC   encoding CD4 was linked to IgG1 DNA at the Esp site upstream of the hinge
CC   region (see AAA50661). Fusion protein CD4Egamma1 and a nucleic acid
CC   encoding it are claimed. Also claimed are a vector comprising the nucleic
CC   acid, and a method of producing the fusion protein in secreted form using
CC   a transformed host cell. The fusion protein may further comprise a
CC   therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
CC   can be administered to an animal (including humans) for treatment of HIV
CC   or SIV infection, and can also be used in assays for HIV or SIV, imaging
CC   and tissue stains. IgG1 fusion proteins such as CD4Egamma1 provide both
CC   complement-mediated and cell-mediated immunity
XX
SQ   Sequence 631 AA;

  Query Match             87.5%;  Score 2034.5;  DB 3;  Length 631;
  Best Local Similarity   92.8%;  Pred. No. 1.5e-130;
  Matches  400;  Conservative    1;  Mismatches   21;  Indels    9;  Gaps    1;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------PRASALPAPPTGSALP 411
              ||||||||||||||||||||||||||||||||||          :    |  |    |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEEPKSCDKTHTCPPCPAPELLG 420

Qy        412 DPQTASALPDP 422
              |    | |
Db        421 GPSVFLFPPKP 431
```

RESULT 34
AAY51079
ID AAY51079 standard; protein; 631 AA.
XX
AC AAY51079;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human fusion protein CD4Egamma1.
XX
KW Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;
KW anti-human immunodeficiency virus; CD4Egamma1.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US6004781-A.
XX
PD 21-DEC-1999.
XX
PF 04-FEB-1994; 94US-00191708.
XX
PR 22-JAN-1988; 88US-00147351.
PR 23-JAN-1989; 89US-00299596.
PR 09-JUN-1992; 92US-00896781.
PR 12-APR-1993; 93US-00057952.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-085792/07.
DR N-PSDB; AAZ44062.
XX
PT Fusion protein useful for the treatment of human immunodeficiency virus.
XX
PS Example 1; Col 29-42; 39pp; English.
XX
CC This invention describes a novel nucleic acid (I) encoding a fusion
CC protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
CC and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
CC chain (III). The products pf the invention have anti-human
CC immunodeficiency virus (HIV) activity and are capable of binding to
CC gp120. The fusion protein is useful for treating human immunodeficiency
CC virus (HIV) or simian immunodeficiency virus (SIV). This sequence
CC represents the fusion protein CD4Egamma1 which is constructed from CD4
CC linked to human IgG1 upstream of the hinge region
XX
SQ Sequence 631 AA;

Query Match 87.5%; Score 2034.5; DB 3; Length 631;
Best Local Similarity 92.8%; Pred. No. 1.5e-130;
Matches 400; Conservative 1; Mismatches 21; Indels 9; Gaps 1;

```
Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------PRASALPAPPTGSALP 411
              ||||||||||||||||||||||||||||||||||          :    |  |    |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEEPKSCDKTHTCPPCPAPELLG 420

Qy        412 DPQTASALPDP 422
              |    | |
Db        421 GPSVFLFPPKP 431
```

RESULT 35
AAY59169
ID AAY59169 standard; protein; 631 AA.
XX
AC AAY59169;
XX
DT 14-MAR-2000 (first entry)
XX
DE CD4-Ig fusion protein CD4Egamma1.
XX
KW HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
KW secreted protein; SIV infection; medicament.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN CA1340741-C.
XX
PD 14-SEP-1999.
XX
PF 20-JAN-1989; 89CA-00588749.
XX
PR 20-JAN-1989; 89CA-00588749.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-063015/06.
DR N-PSDB; AAZ48202.
XX
PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
PT the treatment of HIV or simian immunodeficiency virus infections.
XX
PS Example 1; Page 37-46; 89pp; English.
XX
CC The invention provides a fusion gene encoding a fusion protein that
CC comprises an extracellular CD4 DNA sequence or its fragment which binds
CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
CC sequence of an Ig heavy or light chain, where the DNA sequence encoding

```
CC   the variable region has been replaced with the DNA sequence which encodes
CC   extracellular CD4 or its gp120 binding fragment. The fusion protein is
CC   capable of being secreted. The fusion proteins are useful for treating
CC   HIV or SIV infections in animals, preferably humans. They are also useful
CC   for producing medicaments which can be used for treating HIV or SIV
CC   infections in humans. The present sequence represents the fusion protein
CC   CD4Egamma1 where the CD4 is linked to human IgG1 at the Esp site upstream
CC   of the hinge region
XX
SQ   Sequence 631 AA;

  Query Match             87.5%;  Score 2034.5;  DB 3;  Length 631;
  Best Local Similarity   92.8%;  Pred. No. 1.5e-130;
  Matches  400;  Conservative    1;  Mismatches   21;  Indels    9;  Gaps    1;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------PRASALPAPPTGSALP 411
              ||||||||||||||||||||||||||||||||||||         :   |  |    |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEEPKSCDKTHTCPPCPAPELLG 420

Qy        412 DPQTASALPDP 422
              |     | |
Db        421 GPSVFLFPPKP 431


RESULT 36
AAR78673
ID   AAR78673 standard; protein; 398 AA.
XX
AC   AAR78673;
XX
DT   12-APR-1996  (first entry)
XX
DE   CD4 domains D1-D4.
XX
KW   Chimeric receptor; CD4; T-cell receptor; HIV; cytolysis;
KW   human immunodeficiency virus; adoptive immunotherapy.
XX
OS   Homo sapiens.
XX
PN   WO9521528-A1.
XX
PD   17-AUG-1995.
XX
PF   12-JAN-1995;   95WO-US000454.
XX
PR   14-FEB-1994;   94US-00195395.
PR   02-AUG-1994;   94US-00284391.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Banapour B,  Romeo C,  Kolanus W;
XX
DR   WPI; 1995-292893/38.
DR   N-PSDB; AAQ96103.
XX
PT   Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
PT   cells.
XX
PS   Example 10; Fig 23; 118pp; English.
XX
CC   Extracellular domains D1-D4 (AAR78673) or D1-D2 (AAR78674) of human CD4
CC   are used in the construction of chimeric receptors utilised in the
CC   targeted cytolysis of cells expressing HIV envelope proteins on their
CC   surface. The chimeric receptors comprise the extracellular domain (pref.
CC   amino acids 1-394 or 1-200) of CD4 linked to an intracellular portion,
CC   e.g. of T-cell receptor protein zeta
XX
SQ   Sequence 398 AA;

  Query Match             87.5%;  Score 2034;  DB 2;  Length 398;
  Best Local Similarity   99.7%;  Pred. No. 9.5e-131;
  Matches  394;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVLMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
              ||||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 37
AAY88329
ID   AAY88329 standard; protein; 458 AA.
XX
AC   AAY88329;
XX
DT   14-JUL-2000  (first entry)
XX
DE   T4 glycoprotein amino acid sequence.
XX
KW   sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;
KW   AIDS; treatment; inhibit; cell to cell spread; infection; fusion.
XX
OS   Mammalia.
XX
PN   US5126433-A.
XX
PD   30-JUN-1992.
XX
PF   23-OCT-1987;   87US-00114244.
```

```
XX
PR   21-AUG-1986;   86US-00898587.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Maddon PJ,  Chess L,  Axel R,  Weiss R,  Littman DR,  Mcdougal JS;
XX
DR   WPI; 2000-348913/30.
XX
PT   Soluble T4 glycoprotein useful for prevention and treatment of acquired
PT   immunodeficiency syndrome and for screening inhibitors of human
PT   immunodeficiency viral binding.
XX
PS   Example; Fig 6; 64pp; English.
XX
CC   This sequence represents the amino acid sequence of glycosylated sT4
CC   glycoprotein. Human immunodeficiency virus (HIV) uses sT4 as a target
CC   receptor on T cells. The invention relates to glycosylated sT4 which
CC   functions by blocking the binding of HIV to T4 target cells, and can be
CC   used for the prophylaxis and treatment of AIDS patients. Administration
CC   of sT4 effectively inhibits the cell to cell spreading of HIV infection
CC   and also the fusion of HIV-infected T4 cells and non-infected T4 cells.
CC   The administration of T4 alleviates several symptoms associated with
CC   AIDS, and prevents the occurrence of new pathological changes. The sT4
CC   glycoprotein is useful for the prophylaxis and treatment of patients with
CC   AIDS. It is also useful as a reagent to identify natural, synthetic or
CC   recombinant molecules which act as therapeutic agents or inhibitors of
CC   T4+ cell interactions and in diagnostic assays for detection T4 proteins
CC   or molecules
XX
SQ   Sequence 458 AA;

  Query Match             87.4%;  Score 2032;  DB 3;  Length 458;
  Best Local Similarity   89.2%;  Pred. No. 1.5e-130;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                  ||    |   || :|||:  ||   | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 38
AAB81502
ID   AAB81502 standard; protein; 458 AA.
XX
AC   AAB81502;
XX
DT   18-JUN-2001  (first entry)
XX
DE   Human CD4 protein.
XX
KW   Human; CD4; CD4 fusion protein; oligomerisation;
KW   receptor-ligand interaction inhibition; surface plasmon resonance; SPR;
KW   T cell receptor binding; MHC binding; carcinoma; autoimmune disease;
KW   multiple sclerosis; human immunodeficiency virus; HIV; diabetes;
KW   rheumatoid arthritis; immune disorder.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .25
FT                   /label= Signal_peptide
FT   Protein         26. .458
FT                   /label= Human_CD4
XX
PN   WO200122084-A2.
XX
PD   29-MAR-2001.
XX
PF   18-SEP-2000; 2000WO-GB003579.
XX
PR   21-SEP-1999;   99GB-00022352.
XX
PA   (AVID-) AVIDEX LTD.
XX
PI   Jakobsen BK;
XX
DR   WPI; 2001-273470/28.
DR   N-PSDB; AAF82582.
XX
PT   Sequential screening of candidate compounds library for those which
PT   inhibit binding of low affinity receptor-ligand interaction having fast
PT   binding kinetics, using interfacial optical assay.
XX
PS   Disclosure; Fig 13; 91pp; English.
XX
CC   The present sequence is human CD4. Human CD4 extracellular domains 1 and
CC   2 were used in the construction of CD4 oligomerisation fusion proteins.
CC   The fusion proteins contain an oligomerisation domain that enables the
CC   proteins to bind to one another to form oligomers. The oligomers may be
CC   used in an invention relating to a method for screening for compounds
CC   with the ability to inhibit a low affinity receptor-ligand interaction.
CC   The method uses an interfacial optical assay, such as surface plasmon
CC   resonance (SPR). The method is useful for screening candidate compounds
CC   for the ability to inhibit interaction between MHC/peptide complex and T
CC   cell receptor, and MHC/peptide complex and CD8 or CD4 co-receptor. The
CC   compounds identified by the above methods which interfere with T cell
CC   receptor binding to a particular HLA type molecule are useful as immune
CC   inhibitors for treating carcinomas, autoimmune diseases such as multiple
CC   sclerosis, human immunodeficiency virus (HIV) infection, rheumatoid
CC   arthritis, Hashimoto's disease, insulin dependent diabetes, Good
CC   pasture's syndrome, uveitis, psoriasis and graft rejection
XX
SQ   Sequence 458 AA;

  Query Match             87.4%;  Score 2032;  DB 4;  Length 458;
  Best Local Similarity   89.2%;  Pred. No. 1.5e-130;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
                |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                  ||     |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 39
ADD25609
ID   ADD25609 standard; protein; 458 AA.
XX
AC   ADD25609;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Binding domain-immunoglobulin fusion protein-associated protein #82.
XX
KW   Binding domain; immunoglobulin; fusion protein; cytostatic;
KW   antiarthritic; immunosuppressive; antidiabetic; antithyroid;
KW   neuroprotective; hinge region; immunoglobulin heavy chain;
KW   CH2 constant region; CH3 constant region; IgG1;
KW   antibody dependent cell-mediated cytotoxicity; ADCC; complement fixation;
KW   malignant condition; B-cell disorder; melanoma; carcinoma; sarcoma;
KW   rheumatoid arthritis; myasthenia gravis; Grave's disease;
KW   type I diabetes mellitus; multiple sclerosis; autoimmune disease.
XX
OS   Unidentified.
XX
PN   US2003118592-A1.
XX
PD   26-JUN-2003.
XX
PF   25-JUL-2002; 2002US-00207655.
XX
PR   17-JAN-2001; 2001US-0367358P.
PR   17-JAN-2002; 2002US-00053530.
PR   03-JUN-2002; 2002US-0385691P.
XX
PA   (GENE-) GENECRAFT INC.
XX
PI   Ledbetter JA,  Hayden-Ledbetter MS,  Thompson PA;
XX
DR   WPI; 2003-801317/75.
XX
PT   New binding domain-immunoglobulin fusion protein, useful for treating a
PT   subject having or suspected of having a malignant condition or a B-cell
PT   disorder, e.g. melanoma, Grave's disease or autoimmune disease.
XX
PS   Disclosure; SEQ ID NO 170; 157pp; English.
XX
CC   The invention relates to a binding domain-immunoglobulin fusion protein
CC   comprising a binding domain polypeptide that is fused to an
CC   immunoglobulin hinge region polypeptide, an immunoglobulin heavy chain
CC   CH2 constant region polypeptide that is fused to the hinge region
CC   polypeptide, and an immunoglobulin heavy chain CH3 constant region
CC   polypeptide that is fused to the CH2 constant region polypeptide. The
CC   hinge region polypeptide comprises: a wild-type human IgG1 immunoglobulin
CC   hinge region polypeptide; a mutated human IgG1 immunoglobulin hinge
CC   region polypeptide, derived from (a) having 3 or more cysteine residues;
CC   where the mutated human IgG1 immunoglobulin hinge region polypeptide
CC   contains 2 cysteine residues, where the first cysteine is not mutated; a
CC   mutated human IgG1 immunoglobulin hinge region polypeptide, derived from
CC   (a) having 3 or more cysteine residues, where the mutated human IgG1
CC   immunoglobulin hinge region polypeptide contains no more than one
CC   cysteine residue; and a mutated human IgG1 immunoglobulin hinge region
CC   polypeptide, derived from (a) having 3 or more cysteine residues; where
CC   the mutated human IgG1 immunoglobulin hinge region polypeptide contains
CC   no cysteine residues. The binding domain-immunoglobulin fusion protein is
CC   capable of at least one immunological activity comprising antibody
CC   dependent cell-mediated cytotoxicity (ADCC) and complement fixation. The
CC   binding domain polypeptide is capable of specifically binding to an
CC   antigen. Also included are an isolated polynucleotide encoding the
CC   binding domain-immunoglobulin fusion protein, a recombinant expression
CC   construct comprising the polynucleotide (operably linked to a promoter),
CC   a host cell transformed or transfected with a recombinant expression
CC   construct, producing the binding domain-immunoglobulin fusion protein, a
CC   pharmaceutical composition comprising the binding domain-immunoglobulin
CC   fusion protein or polynucleotide and a carrier, and treating a subject
CC   having or suspected of having a malignant condition or a B-cell disorder.
CC   The binding domain-immunoglobulin fusion protein is useful for treating a
CC   subject having or suspected of having a malignant condition or a B-cell
CC   disorder, e.g. melanoma, carcinoma or sarcoma, rheumatoid arthritis,
CC   myasthenia gravis, Grave's disease, type I diabetes mellitus, multiple
CC   sclerosis or autoimmune disease. The present sequence is a binding domain
CC   -immunoglobulin fusion protein-associated protein sequence. Note: The
CC   sequence data for this patent formed part of the printed specification
CC   and is also available in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030118592. The authors have not
CC   identified the sequences in the printed specification by their SEQ ID
CC   number therefore none of the sequences can be explicitly identified.
XX
SQ   Sequence 458 AA;

  Query Match             87.4%;  Score 2032;  DB 7;  Length 458;
  Best Local Similarity   89.2%;  Pred. No. 1.5e-130;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                  ||     |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425
```

```
RESULT 40
ADE57489
ID   ADE57489 standard; protein; 458 AA.
XX
AC   ADE57489;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human Protein P01730, SEQ ID NO 3351.
XX
KW   Human; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; P01730.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 458 AA;

  Query Match             87.4%;  Score 2032;  DB 7;  Length 458;
  Best Local Similarity   89.2%;  Pred. No. 1.5e-130;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP------------------------ 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |    || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 41
ADA44807
ID   ADA44807 standard; protein; 473 AA.
XX
AC   ADA44807;
XX
DT   04-DEC-2003  (first entry)
XX
DE   CD4/TCR CD3epsilon chain chimeric protein CD4epsilon15, SEQ ID NO:2.
XX
KW   HIV-1 infection; human immunodeficiency virus-1; CD4+ cell; chimeric CD4;
KW   endoplasmic reticulum; ER retention; envelope protein gp160;
KW   T cell receptor CD3epsilon chain; C-terminal domain; CD4epsilon15;
KW   gene therapy; human; receptor.
XX
OS   Chimeric.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         1. .458
FT                   /label= CD4
FT   Region          459. .473
FT                   /note= "Part of the C-terminal domain of the T cell
FT                   receptor CD3epsilon chain"
XX
PN   WO2003076468-A1.
XX
PD   18-SEP-2003.
XX
PF   14-MAR-2003; 2003WO-ES000120.
XX
PR   14-MAR-2002; 2002ES-00000616.
XX
PA   (CNSJ ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
XX
PI   Alarcon Sanchez BJ,  San Jose Martinez ME,  Zaldivar Notario I;
PI   Gomez Buendia M;
XX
DR   WPI; 2003-779059/73.
```

```
DR   N-PSDB; ADA44806.
XX
PT   Composition for treating or preventing human immune deficiency virus,
PT   comprises CD4 chimeric protein having a protective effect in trans, or
PT   related nucleic acid.
XX
PS   Claim 5; Page 33-35; 43pp; Spanish.
XX
CC   The invention relates to a composition for the treatment or prevention of
CC   human immunodeficiency virus-1 (HIV-1) infection. The composition
CC   comprises CD4+ cells that have been transduced with a vector that encodes
CC   a chimeric CD4 molecule which is capable of being retained in the
CC   endoplasmic reticulum (ER). The invention also encompasses the use of a
CC   soluble protein factor produced by CD4+ cells that have been transduced
CC   with a vector encoding a chimeric CD4 protein; and the use of an
CC   expression system encoding a chimeric CD4 protein. The ER-localised
CC   chimeric CD4 molecule binds to the HIV-1 envelope protein gp160,
CC   resulting in HIV-1 retention in the ER and thereby preventing viral
CC   replication. In a specific embodiment, the chimeric CD4 molecule
CC   comprises CD4 fused to 15 amino acids of the C-terminal domain of the T
CC   cell receptor CD3epsilon chain; this chimeric CD4 molecule is designated
CC   CD4epsilon15 (ADA44807). A known chimeric CD4 of similar structure but
CC   containing only 10 amino acids from CD3epsilon can also be used.
CC   Compositions of the invention have an in trans effect on the replication
CC   of HIV-1, and may be used to treat and prevent HIV-1 infection. The
CC   present sequence represents the chimeric CD4 molecule CD4epsilon15, which
CC   is specifically claimed for use in compositions of the invention.
XX
SQ   Sequence 473 AA;

  Query Match             87.4%;  Score 2032;  DB 7;  Length 473;
  Best Local Similarity   89.2%;  Pred. No. 1.6e-130;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                ||      |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 42
AAR20152
ID   AAR20152 standard; protein; 519 AA.
XX
AC   AAR20152;
XX
DT   25-MAR-2003  (revised)
DT   31-MAR-1992  (first entry)
XX
DE   Human CD4 sequence encoded by PATY.6.
XX
KW   Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
KW   acquired immune deficiency syndrome; AIDS related complex;
KW   T helper lymphocytes.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .25
FT                   /label= signal_sequence
XX
PN   WO9118618-A.
XX
PD   12-DEC-1991.
XX
PF   25-MAY-1990;   90US-00529186.
XX
PR   25-MAY-1990;   90US-00529186.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Fisher RA,  Hession C,  Burkly LC;
XX
DR   WPI; 1992-007200/01.
DR   N-PSDB; AAQ20327.
XX
PT   New immuno-therapeutic human CD4 variants and derivs. - elicit AB
PT   production to HIV gp.120, useful in treating, preventing and diagnosing
PT   AIDS, ARC and HIV infections.
XX
PS   Disclosure; Fig 28; 179pp; English.
XX
CC   The sequence was deduced from the DNA sequence of subclone PATY.6, contg.
CC   DNA coding for the full-length human CD4. The clone was constructed from
CC   plasmids pBG178A and pBG378 (both in US8802940). The DNA can be used to
CC   express recombinant CD4 and analogues for use in diagnosis and treatment
CC   of diseases caused by infective agents whose primary targets are T4+
CC   lymphocytes. See also AAR20148-R20155 and AAR21078. (Updated on 25-MAR-
CC   2003 to correct PA field.)
XX
SQ   Sequence 519 AA;

  Query Match             87.4%;  Score 2032;  DB 2;  Length 519;
  Best Local Similarity   89.2%;  Pred. No. 1.8e-130;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         62 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 121

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 181

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 241

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 301

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 361

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        362 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 421
```

```
Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||||| |
Db        422 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 457

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                ||     |   || :|||:  ||  | |
Db        458 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 486


RESULT 43
AAR06374
ID   AAR06374 standard; protein; 400 AA.
XX
AC   AAR06374;
XX
DT   31-OCT-2002  (revised)
DT   20-DEC-1990  (first entry)
XX
DE   Truncated form of soluble T4 encoded by pBG381.
XX
KW   plasmid pBG381; soluble T4 protein; AIDS; ARC; HIV.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= secretory signal
FT                   /note= "hydrophobic"
FT   Region          24. .117
FT                   /label= extracellular
FT                   /note= "homology to V-regions"
FT   Region          118. .132
FT                   /label= extracellular
FT                   /note= "homology to J-regions"
FT   Region          133. .397
FT                   /label= extracellular
FT                   /note= "glycosylated"
XX
PN   WO9008198-A.
XX
PD   26-JUL-1990.
XX
PF   18-JAN-1989;   89US-00300096.
XX
PR   18-JAN-1989;   89US-00300096.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Letvin NA;
XX
DR   WPI; 1990-254040/33.
DR   N-PSDB; AAQ05608.
XX
PT   Treating or preventing AIDS, ARC or HIV infection - by administering an
PT   immunologically effective amt. of soluble T4 protein.
XX
PS   Disclosure; Fig 2; 121pp; English.
XX
CC   T4-encoding plasmid pBG381 was used to transform Chinese Hamster Ovary
CC   cells for the production of truncated T4. Soluble T4 is produced by
CC   virtue of the removal of the transmembrane and cytoplasmic domains. The
CC   soluble forms may be modified to increase their immunogenicity by
CC   addition of an adjuvant such as incomplete Freund's adjuvant. The T4
CC   interferes with HIV/T4 interaction and elicits anti-soluble T4 antibody
CC   production. See also AAQ05607. (Updated on 31-OCT-2002 to add missing OS
CC   field.)
XX
SQ   Sequence 400 AA;

  Query Match             87.4%;  Score 2031;  DB 2;  Length 400;
  Best Local Similarity   99.2%;  Pred. No. 1.5e-130;
  Matches  395;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRA 398
              |||||||||||||||||||||||||||||||| | |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMA 398


RESULT 44
AAY51081
ID   AAY51081 standard; protein; 481 AA.
XX
AC   AAY51081;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Human fusion protein CD4Pmu.
XX
KW   Fusion protein; human; CD4; IgM; immunoglobulin; gp120;
KW   anti-human immunodeficiency virus; CD4Pmu.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   US6004781-A.
XX
PD   21-DEC-1999.
XX
PF   04-FEB-1994;   94US-00191708.
XX
PR   22-JAN-1988;   88US-00147351.
PR   23-JAN-1989;   89US-00299596.
PR   09-JUN-1992;   92US-00896781.
PR   12-APR-1993;   93US-00057952.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-085792/07.
DR   N-PSDB; AAZ44064.
XX
PT   Fusion protein useful for the treatment of human immunodeficiency virus.
XX
PS   Example 1; Col 49-58; 39pp; English.
XX
CC   This invention describes a novel nucleic acid (I) encoding a fusion
CC   protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
CC   and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
CC   chain (III). The products pf the invention have anti-human
CC   immunodeficiency virus (HIV) activity and are capable of binding to
```

```
CC   gp120. The fusion protein is useful for treating human immunodeficiency
CC   virus (HIV) or simian immunodeficiency virus (SIV). This sequence
CC   represents the fusion protein CD4Pmu which is constructed from CD4 linked
CC   to human IgM upstream of the CH2 region
XX
SQ   Sequence 481 AA;

  Query Match             87.4%;  Score 2031;  DB 3;  Length 481;
  Best Local Similarity   93.0%;  Pred. No. 1.9e-130;
  Matches  399;  Conservative    4;  Mismatches   10;  Indels   16;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||| ||||||||||||||||              ||:  : ||
Db        361 LNPEAGMWQCLLSDSGQBLLESNIKVLPTWSTPVH--------------ADPEVIAELP 405

Qy        421 DPPAASALP 429
              |  :  :|
Db        406 -PKVSVFVP 413


RESULT 45
AAR04032
ID   AAR04032 standard; protein; 2037 AA.
XX
AC   AAR04032;
XX
DT   25-MAR-2003  (revised)
DT   31-OCT-2002  (revised)
DT   29-MAY-1990  (first entry)
XX
DE   Full length T4 encoded by plasmid pBG381.
XX
KW   Soluble T4; pBG381; anti-retroviral agent; AIDS; ARC; HIV; AZT.
XX
OS   Synthetic.
XX
PN   WO8911860-A.
XX
PD   14-DEC-1989.
XX
PF   08-JUN-1989;   89WO-US002453.
XX
PR   10-JUN-1988;   88US-00204645.
PR   20-APR-1989;   89US-00341080.
XX
PA   (BIOJ ) BIOGEN NV INC.
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (BIOJ ) BIOGEN INC.
PA   (BIOJ ) BIOGEN INC.
XX
PI   Fisher RA,  Schooley RT,  Hirsch MS,  Johnson VA,  Walker BD;
XX
DR   WPI; 1990-007302/01.
DR   N-PSDB; AAQ03006.
XX
PT   Combinations of soluble T4 protein and anti-retro:viral agent - having
PT   synergistic activity in treatment and prevention of AIDS, arc and HIV
PT   infection.
XX
PS   Disclosure; Fig 2; 100pp; English.
XX
CC   X = stop codon. The sequence was deduced from the cDNA insert of pBG183.
CC   Soluble T4 constructs may be produced by truncating this sequence to give
CC   fragments from position 400 to 799, removing the transmembrane and
CC   intracytoplasmic domains whilst retaining the extracellular region
CC   responsible for HIV binding. The sol. T4 is combined with an anti-viral
CC   agent such as AZT. See also AAQ03005. (Updated on 31-OCT-2002 to add
CC   missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 2037 AA;

  Query Match             87.4%;  Score 2031;  DB 2;  Length 2037;
  Best Local Similarity   99.2%;  Pred. No. 1.1e-129;
  Matches  395;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 462

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 522

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 582

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 642

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 702

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        703 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 762

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRA 398
              ||||||||||||||||||||||||||||||||| | |
Db        763 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMA 800


RESULT 46
AAR07641
ID   AAR07641 standard; protein; 2050 AA.
XX
AC   AAR07641;
XX
DT   31-OCT-2002  (revised)
DT   20-DEC-1990  (first entry)
XX
DE   Deduced sequence of pBG381 comprising truncated T4 glycoprotein.
XX
KW   plasmid pBG381; soluble T4 protein; AIDS; ARC; HIV.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         403. .803
FT                   /label= truncated soluble T4 glycoprotein
```

```
XX
PN   WO9008198-A.
XX
PD   26-JUL-1990.
XX
PF   18-JAN-1989;   89US-00300096.
XX
PR   18-JAN-1989;   89US-00300096.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Letvin NA;
XX
DR   WPI; 1990-254040/33.
DR   N-PSDB; AAQ05608.
XX
PT   Treating or preventing AIDS, ARC or HIV infection - by administering an
PT   immunologically effective amt. of soluble T4 protein.
XX
PS   Disclosure; Fig 2; 121pp; English.
XX
CC   Entire sequence translation of plasmid pBG381 used to transform Chinese
CC   Hamster Ovary cells for the production of soluble truncated T4.
CC   Transmembrane and cytoplasmic domain-encoding regions are deleted from
CC   the T4 CDS to encode a truncated protein. The soluble forms may be
CC   modified to increase their immunogenicity by addition of an adjuvant such
CC   as incomplete Freund's adjuvant. The T4 interferes with HIV/T4
CC   interaction and elicits anti-soluble T4 antibody prodn. See also
CC   AAQ05607. (Updated on 31-OCT-2002 to add missing OS field.)
XX
SQ   Sequence 2050 AA;

  Query Match             87.4%;  Score 2031;  DB 2;  Length 2050;
  Best Local Similarity   99.2%;  Pred. No. 1.1e-129;
  Matches  395;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 462

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 522

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 582

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 642

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        643 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 702

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        703 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 762

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRA 398
              ||||||||||||||||||||||||||||||||| | ||
Db        763 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMA 800


RESULT 47
AAR13491
ID   AAR13491 standard; protein; 458 AA.
XX
AC   AAR13491;
XX
DT   25-MAR-2003  (revised)
DT   30-OCT-1991  (first entry)
XX
DE   Human CD4 encoded by pJOD.sCD4.Y187.SnaB1 and p170.2.
XX
KW   C4bp; gp120; HIV; T lymphocyte; fusion protein.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .25
FT                   /label= signal_peptide
FT   Domain          26. .132
FT                   /label= Ig-related
FT                   /note= "extracellular"
FT   Disulfide-bond  41. .109
FT   Domain          133. .202
FT                   /label= Ig-related
FT                   /note= "extracellular"
FT   Disulfide-bond  155. .184
FT   Domain          203. .318
FT                   /label= Ig-related
FT                   /note= "extracellular"
FT   Domain          319. .395
FT                   /label= Ig-related
FT                   /note= "extracellular"
FT   Disulfide-bond  328. .370
FT   Region          396. .416
FT                   /label= transmembrane
FT   Domain          417. .456
FT                   /label= cytoplasmic
XX
PN   WO9111461-A.
XX
PD   08-AUG-1991.
XX
PF   26-JAN-1990;   90US-00470888.
XX
PR   26-JAN-1990;   90US-00470888.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Pasek MP,  Winkler G,  Liu TR;
XX
DR   WPI; 1991-252613/34.
DR   N-PSDB; AAQ13243.
XX
PT   New C4 binding protein fusion proteins and DNA encoding them - comprise
PT   assemblies of C4bp monomers linked to functional moiety, e.g. AZT, useful
PT   as delivery vehicles in diagnosis and therapy.
XX
PS   Example 3; Fig 3; 105pp; English.
XX
CC   This is the preferred CD4 sequence for use in the construction of fusion
CC   proteins with C4-binding protein. Truncated, soluble versions of CD4 can
CC   also be used. The C4bp-CD4 fusion protein may be useful to target AZT or
CC   similar anti-retroviral agent to HIV-infected cells. See AAQ13242-51.
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 458 AA;

  Query Match             87.3%;  Score 2030;  DB 2;  Length 458;
  Best Local Similarity   89.2%;  Pred. No. 2.1e-130;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITSDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                  ||      |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 48
AAP93011
ID   AAP93011 standard; protein; 481 AA.
XX
AC   AAP93011;
XX
DT   25-MAR-2003  (revised)
DT   03-AUG-1992  (first entry)
XX
DE   Genetic construct which encodes CD4 linked to human IgM at the Pst site
DE   upstream of the CH2 region (fusion protein CD4Pmu).
XX
KW   Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
KW   diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
XX
OS   Homo sapiens.
XX
PN   EP325262-A.
XX
PD   26-JUL-1989.
XX
PF   20-JAN-1989;   89EP-00100913.
XX
PR   22-JAN-1988;   88US-00147351.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 1989-214472/30.
DR   N-PSDB; AAN90359.
XX
PT   Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PT   infections or detecting HIV or SIV in sample.
XX
PS   Example; Table 4, Page 41-47; 68pp; English.
XX
CC   The fusion protein genes of the invention pref. comprises cDNA sequences
CC   which encode CD4 or a fragment which binds gp120 ligated to an expression
CC   plasmid which encodes an antibody in which the variable region of the
CC   gene has been deleted (see Wo87-02671). The CD4 portion of the fusion
CC   protein may comprise the complete CD4 sequence, the 370 AA extracellular
CC   region and the membrane spanning domain, or the extracellular region. The
CC   Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are
CC   specifically claimed: fusion proteins CD4Hlambda1, CD4Mmu, CD4Pmu,
CC   CD4Elambda1, and CD4Mmu (No. 67608), pCD4Plambda (No. 67609) and
CC   pCD4Elambda1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 481 AA;
```

```
  Query Match             87.3%;  Score 2030;  DB 1;  Length 481;
  Best Local Similarity   93.0%;  Pred. No. 2.2e-130;
  Matches  399;  Conservative    4;  Mismatches   10;  Indels   16;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEAQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------------ADPEVIAELP 405

Qy        421 DPPAASALP 429
              | : :|
Db        406 -PKVSVFVP 413


RESULT 49
AAP93012
ID   AAP93012 standard; protein; 614 AA.
XX
AC   AAP93012;
XX
DT   25-MAR-2003  (revised)
DT   03-AUG-1992  (first entry)
XX
DE   Genetic construct which encodes CD4 linked to human IgG1 at the Ban1 site
DE   downstream from the hinge region (fusion protein CD4Blambda1).
XX
KW   Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
KW   diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
XX
OS   Homo sapiens.
XX
PN   EP325262-A.
XX
PD   26-JUL-1989.
XX
PF   20-JAN-1989;   89EP-00100913.
XX
PR   22-JAN-1988;   88US-00147351.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 1989-214472/30.
DR   N-PSDB; AAN90360.
XX
PT   Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PT   infections or detecting HIV or SIV in sample.
XX
PS   Example; Table 5, Page 48-55; 68pp; English.
```

```
XX
CC   The fusion protein genes of the invention pref. comprises cDNA sequences
CC   which encode CD4 or a fragment which binds gp120 ligated to an expression
CC   plasmid which encodes an antibody in which the variable region of the
CC   gene has been deleted (see Wo87-02671). The CD4 portion of the fusion
CC   protein may comprise the complete CD4 sequence, the 370 AA extracellular
CC   region and the membrane spanning domain, or the extracellular region. The
CC   Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are
CC   specifically claimed: fusion proteins CD4Hlambda1, CD4Mmu, CD4Pmu,
CC   CD4Elambda1, and CD4Mmu (No. 67608), pCD4Plambda (No. 67609) and
CC   pCD4Elambda1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 614 AA;

  Query Match             87.1%;  Score 2026;  DB 1;  Length 614;
  Best Local Similarity   96.1%;  Pred. No. 5.7e-130;
  Matches  397;  Conservative    1;  Mismatches   13;  Indels    2;  Gaps    1;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQNKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PRASALPAPPTGSALP 411
              ||||||||||||| |||||||||||||||||||  | |  |   |:    |
Db        361 LNPEAGMWQCLLSDSAQVLLESNIKVLPTWSTPVHADPEAPELLGGPSVFLFP 413


RESULT 50
AAG79087
ID   AAG79087 standard; protein; 458 AA.
XX
AC   AAG79087;
XX
DT   10-DEC-2001  (first entry)
XX
DE   Amino acid sequence of human CD4 protein.
XX
KW   Human; receptor; DC-SIGN; dendritic cell; T lymphocyte; HIV; gp120;
KW   C-type lectin; ICAM3; HIV entry; T cell; macrophage; HIV infection; CD4.
XX
OS   Homo sapiens.
XX
PN   WO200164752-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US006322.
XX
PR   02-MAR-2000; 2000US-00517605.
XX
PA   (UYNY ) UNIV NEW YORK STATE.
PA   (UYNI-) UNIV NIJMEGEN.
XX
PI   Littman DR,  Kwon D,  Van Kooyk Y,  Geijtenbeek T;
XX
DR   WPI; 2001-602565/68.
XX
PT   An antibody for the treatment or prevention of HIV-infection comprises a
PT   gp120 portion which binds to DC-SIGN or is exposed upon gp120 binding of
PT   DC-SIGN due to concomitant conformational change.
XX
PS   Disclosure; Page 115-116; 131pp; English.
XX
CC   The specification describes an antibody which is specific for an
CC   antigenic fragment of gp120. This antigenic fragment binds to DC-SIGN or
CC   is exposed upon gp120 binding of DC-SIGN due to concomitant
CC   conformational change. DC-SIGN is a receptor that is specifically
CC   expressed on dendritic cells and facilitates infection of T lymphocytes
CC   with HIV. DC-SIGN is identical to a HIV-1 gp120-binding C-type lectin. DC
CC   -SIGN binds ICAM-3 (which is expressed constitutively on T lymphocytes)
CC   with high affinity. The antibody of the invention inhibits the trans
CC   enhancement of HIV entry into a T cell or macrophage facilitated by
CC   dendritic cells. The antibody is useful to treat or prevent HIV
CC   infection. The present sequence represents a human CD4 protein
XX
SQ   Sequence 458 AA;

  Query Match             87.1%;  Score 2024;  DB 4;  Length 458;
  Best Local Similarity   89.0%;  Pred. No. 5.4e-130;
  Matches  404;  Conservative    2;  Mismatches   18;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKWVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 51
AAP93506
ID   AAP93506 standard; protein; 394 AA.
XX
AC   AAP93506;
XX
DT   25-MAR-2003  (revised)
DT   02-JUN-1990  (first entry)
XX
DE   Derived sequence of soluble T4 lymphocyte surface protein (sT4).
XX
KW   Soluble T4 lymphocyte surface protein; sT4; AIDS therapy; AIDS diagnosis.
```

```
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         26. .394
FT   Misc-difference 26. .26
FT                   /note= "When sequence was determined by amino acid
FT                   sequencing, this residue was Lys."
FT   Region          27. .45
FT                   /note= "These residues are identical to those determined
FT                   by amino acid sequencing"
XX
PN   EP313377-A.
XX
PD   26-APR-1989.
XX
PF   21-OCT-1988;   88EP-00309907.
XX
PR   23-OCT-1987;   87US-00112800.
XX
PA   (SMIK ) SMITHKLINE BECKMAN CORP.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Deen KC,  Folenawass GM,  Inacker RH,  Sweet RW;
XX
DR   WPI; 1989-124209/17.
DR   N-PSDB; AAN90763.
XX
PT   Purifying soluble recombinant T4 lymphocyte surface protein - from cell
PT   culture by adsorption on cation exchanger, elution and treatment with
PT   anion exchanger.
XX
PS   Disclosure; Fig 1; 13pp; English.
XX
CC   The coding sequenc is derived from the published sequence of sT4. sT4 is
CC   useful in the prevention and treatment of AIDS by inhibiting spread of
CC   the virus. It can also be used as an inhibitor of T4+ cell function, as a
CC   reagent for identifying inhibitors of T4+ cell interaction and to produce
CC   diagnostic monoclonal antibodies. (Updated on 25-MAR-2003 to correct PA
CC   field.) (Updated on 25-MAR-2003 to correct PI field.)
XX
SQ   Sequence 394 AA;

  Query Match             86.8%;  Score 2018;  DB 1;  Length 394;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-129;
  Matches  392;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLVLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394


RESULT 52
AAP91922
ID   AAP91922 standard; protein; 402 AA.
XX
AC   AAP91922;
XX
DT   25-MAR-2003  (revised)
DT   31-OCT-2002  (revised)
DT   14-MAY-1990  (first entry)
XX
DE   Sequence of a secreted form of the CD4 adhesion (CD4T) polypeptide.
XX
KW   CD4 variants; CD4T; gp120; plasmid pRKCD4; HIV-1; HTLV-IIIB.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 25. .26
FT                   /note= "signal processing site"
FT   Misc-difference 366
FT                   /note= "other forms of CD4T terminate here"
FT   Misc-difference 368
FT                   /note= "other forms of CD4T terminate here"
XX
PN   EP314317-A.
XX
PD   03-MAY-1989.
XX
PF   03-OCT-1988;   88EP-00309194.
XX
PR   02-OCT-1987;   87US-00104329.
PR   28-SEP-1988;   88US-00250785.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Capon DJ,  Gregrory TJ;
XX
DR   WPI; 1989-131855/18.
DR   N-PSDB; AAN90777.
XX
PT   Compsns. contg. adheson variants - useful in therapy and diagnostics,
PT   e.g. CD4 variants which are therapeutically useful for treating human
PT   immuno-deficiency virus.
XX
PS   Disclosure; Fig 1a-1c; 36pp; English.
XX
CC   It may be capable of binding gp120. It may be fused with an
CC   immunoglobulin constant domain, human transferrin, apolipoprotein,
CC   albumin, ricin A chain or diphtheria toxin A. It may be used for
CC   antiviral of immunomodulatory therapy particularly in treatment of HIV
CC   infection. It may have variants by insertion, substitution of deletion in
CC   non-functional regions. (Updated on 31-OCT-2002 to add missing OS field.)
CC   (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
CC   correct PI field.)
XX
SQ   Sequence 402 AA;

  Query Match             86.8%;  Score 2017;  DB 1;  Length 402;
  Best Local Similarity   99.7%;  Pred. No. 1.4e-129;
  Matches  392;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTP 393
                |||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTP 393


RESULT 53
AAP94757
ID   AAP94757 standard; protein; 402 AA.
XX
AC   AAP94757;
XX
DT   25-MAR-2003  (revised)
DT   03-OCT-2002  (revised)
DT   28-JAN-1991  (first entry)
XX
DE   Sequence of a secreted form of the CD4 adheson.
XX
KW   HIV; antiviral; therapy; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .25
FT                   /note= "signal"
FT   Protein         26. .402
XX
PN   WO8902922-A.
XX
PD   06-APR-1989.
XX
PF   03-OCT-1988;   88WO-US003414.
XX
PR   02-OCT-1987;   87US-00104329.
PR   28-SEP-1988;   88US-00250785.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Capon DJ,  Gregory TJ;
XX
DR   WPI; 1989-114397/15.
DR   N-PSDB; AAN90734.
XX
PT   New nucleic acid sequences encoding adheson, esp. CD 4, variants -
PT   partic. with trans-membrane domain inactivated or fused to other peptide,
PT   useful esp. for treating HIV infections.
XX
PS   Disclosure; Fig 1a-1c; 78pp; English.
XX
CC   The patent claims a nucleic acid encoding an aa sequence variant of an
CC   adheson, which is pref. a CD4 polypeptide variant modified such that its
CC   transmembrane domain has been inactivated, either deleted or replaced by
CC   a sequence of hydrophilic hydropathy profile. The aa sequence variant of
CC   an adheson may also be a fusion of CD4 with a 2nd polypeptide esp. one
CC   contg. a non-CD4 epitope; a signal sequence; a cpd. able to elicit a
CC   humoral immune response (viral polypeptide or allergen); or a human
CC   plasma protein of long plasma half-life. CD4 fusion proteins can have
CC   antiviral and immunomodulatory activity and are esp. useful for treating
CC   HIV infections regardless of genetic variation within the virus. They and
CC   antibodies raised against them can also be used diagnostically for
CC   assaying adhesons and their ligands. (Updated on 03-OCT-2002 to add
CC   missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 402 AA;

  Query Match             86.8%;  Score 2017;  DB 1;  Length 402;
  Best Local Similarity   99.7%;  Pred. No. 1.4e-129;
  Matches  392;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTP 393
              |||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTP 393


RESULT 54
AAY88328
ID   AAY88328 standard; protein; 394 AA.
XX
AC   AAY88328;
XX
DT   14-JUL-2000  (first entry)
XX
DE   T4 glycoprotein amino acid sequence.
XX
KW   sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;
KW   AIDS; treatment; inhibit; cell to cell spread; infection; fusion.
XX
OS   Mammalia.
XX
PN   US5126433-A.
XX
PD   30-JUN-1992.
XX
PF   23-OCT-1987;   87US-00114244.
XX
PR   21-AUG-1986;   86US-00898587.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Maddon PJ,  Chess L,  Axel R,  Weiss R,  Littman DR,  Mcdougal JS;
XX
DR   WPI; 2000-348913/30.
DR   N-PSDB; AAA10906.
XX
PT   Soluble T4 glycoprotein useful for prevention and treatment of acquired
PT   immunodeficiency syndrome and for screening inhibitors of human
PT   immunodeficiency viral binding.
```

```
XX
PS   Disclosure; Col 11-16; 64pp; English.
XX
CC   This sequence represents the full length amino acid sequence of
CC   glycosylated sT4 glycoprotein. Human immunodeficiency virus (HIV) uses
CC   sT4 as a target receptor on T cells. The invention relates to
CC   glycosylated sT4 which functions by blocking the binding of HIV to T4
CC   target cells, and can be used for the prophylaxis and treatment of AIDS
CC   patients. Administration of sT4 effectively inhibits the cell to cell
CC   spreading of HIV infection and also the fusion of HIV-infected T4 cells
CC   and non-infected T4 cells. The administration of T4 alleviates several
CC   symptoms associated with AIDS, and prevents the occurrence of new
CC   pathological changes. The sT4 glycoprotein is useful for the prophylaxis
CC   and treatment of patients with AIDS. It is also useful as a reagent to
CC   identify natural, synthetic or recombinant molecules which act as
CC   therapeutic agents or inhibitors of T4+ cell interactions and in
CC   diagnostic assays for detection T4 proteins or molecules
XX
SQ   Sequence 394 AA;

  Query Match             86.7%;  Score 2015;  DB 3;  Length 394;
  Best Local Similarity   99.5%;  Pred. No. 1.9e-129;
  Matches  392;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394
              |||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394


RESULT 55
ADE65841
ID   ADE65841 standard; protein; 458 AA.
XX
AC   ADE65841;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human CD4 receptor.
XX
KW   Human; CD4 receptor; receptor; protein-protein interaction;
KW   protein array; PDZ domain; drug target screening.
XX
OS   Homo sapiens.
XX
PN   US2003170723-A1.
XX
PD   11-SEP-2003.
XX
PF   06-MAR-2002; 2002US-00092138.
XX
PR   06-MAR-2002; 2002US-00092138.
XX
PA   (SATO/) SATO T.
XX
PI   Sato T;
XX
DR   WPI; 2003-852032/79.
XX
PT   Preparing a protein array useful for screening drug targets comprises
PT   depositing an array of a first protein on substrate, and applying a
PT   second protein comprising an amino acid sequence that binds to a domain
PT   of the first protein.
XX
PS   Disclosure; SEQ ID NO 25; 60pp; English.
XX
CC   The invention relates to a method for preparing a protein array based on
CC   protein-protein interaction, by depositing an array of a first protein
CC   comprising a PDZ domain on a substrate, and applying a second protein
CC   comprising an amino acid sequence that binds to the PDZ domain of the
CC   first protein. The method is useful for preparing protein arrays based on
CC   biochemical protein-protein interactions. Arrays produced by this method
CC   are useful for screening drug targets. This sequence represents the human
CC   CD4 receptor, used in the method of the invention.
XX
SQ   Sequence 458 AA;

  Query Match             86.6%;  Score 2014;  DB 7;  Length 458;
  Best Local Similarity   88.5%;  Pred. No. 2.6e-129;
  Matches  402;  Conservative    3;  Mismatches   19;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTITLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||  |||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQENVLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 56
AAY39825
ID   AAY39825 standard; protein; 394 AA.
XX
AC   AAY39825;
XX
DT   03-DEC-1999  (first entry)
XX
DE   Soluble human T4 protein.
```

```
XX
KW   Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;
KW   vaccine; immunisation; therapy.
XX
OS   Homo sapiens.
XX
PN   US5958678-A.
XX
PD   28-SEP-1999.
XX
PF   12-DEC-1994;   94US-00354452.
XX
PR   21-AUG-1986;   86US-00898587.
PR   11-JUN-1991;   91US-00713564.
PR   06-JUL-1992;   92US-00909021.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Mcdougal JS,  Weiss R,  Axel R,  Littman DR,  Maddon PJ,  Chess L;
XX
DR   WPI; 1999-561025/47.
DR   N-PSDB; AAZ20694.
XX
PT   Human T4 protein inhibits HIV binding to T4 cells, useful for treating
PT   AIDS.
XX
PS   Disclosure; Col 13-16; 58pp; English.
XX
CC   This sequence represents the soluble human T4 protein of the invention.
CC   The soluble human T4 protein blocks the binding of HIV to T4+ cells and
CC   is therefore useful for the treatment of AIDS. Monoclonal antibodies
CC   against the T4 protein may be used as vaccines for immunising subjects
CC   against AIDS
XX
SQ   Sequence 394 AA;

  Query Match             86.5%;  Score 2012;  DB 2;  Length 394;
  Best Local Similarity   99.2%;  Pred. No. 3e-129;
  Matches  391;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVDFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394


RESULT 57
AAR06373
ID   AAR06373 standard; protein; 458 AA.
XX
AC   AAR06373;
XX
DT   31-OCT-2002  (revised)
DT   20-DEC-1990  (first entry)
XX
DE   T4 encoded by plasmid p170-2.
XX
KW   plasmid p170-2; soluble T4 protein; AIDS; ARC; HIV.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .23
FT                   /label= hydrophobic/secretory signal
FT   Region          24. .117
FT                   /label= extracellular
FT                   /note= "homology to V-regions"
FT   Region          118. .132
FT                   /label= extracellular
FT                   /note= "homology to J-regions"
FT   Region          133. .397
FT                   /label= extracellular
FT                   /note= "glycosylated region"
FT   Region          398. .418
FT                   /label= transmembrane sequence
FT                   /note= "hydrophobic"
FT   Region          419. .458
FT                   /label= intracytoplasmic
FT                   /note= "very hydrophilic"
XX
PN   WO9008198-A.
XX
PD   26-JUL-1990.
XX
PF   18-JAN-1989;   89US-00300096.
XX
PR   18-JAN-1989;   89US-00300096.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Letvin NA;
XX
DR   WPI; 1990-254040/33.
DR   N-PSDB; AAQ05607.
XX
PT   Treating or preventing AIDS, ARC or HIV infection - by administering an
PT   immunologically effective amt. of soluble T4 protein.
XX
PS   Disclosure; Fig 1; 121pp; English.
XX
CC   Soluble T4 can be produced by truncating the CDS to remove the
CC   transmembrane and cytoplasmic domains. The soluble forms may be modified
CC   to increase their immunogenicity by addition of an adjuvant such as
CC   incomplete Freund's adjuvant. The T4 interferes with HIV/T4 interaction
CC   and elicits anti-soluble T4 antibody production. Soluble T4 proteins
CC   include the following polypeptide fragments: amino acids 1-385, 24-385,
CC   Met-24-385, 24-397, 1-400 and Met-24-400. See also AAQ05608. (Updated on
CC   31-OCT-2002 to add missing OS field.)
XX
SQ   Sequence 458 AA;

  Query Match             86.5%;  Score 2010;  DB 2;  Length 458;
  Best Local Similarity   88.8%;  Pred. No. 4.9e-129;
  Matches  403;  Conservative    2;  Mismatches   19;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLRDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
```

```
Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITSDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||     |    || :|||:  ||   | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 58
AAP94703
ID   AAP94703 standard; protein; 524 AA.
XX
AC   AAP94703;
XX
DT   25-MAR-2003  (revised)
DT   22-MAR-1991  (first entry)
XX
DE   Sequence encoded by T4 lymphocyte cDNA obtained from PBL clone lambda-203
DE   -4.
XX
KW   HIV; soluble T4; immunotherapeutic; prophylactic; diagnostic; AIDS; ARC.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 67
FT                   /note= "AA DESIGNATED NUMBER -23"
FT   Misc-difference 90
FT                   /note= "AA DESIGNATED NUMBER 1"
FT   Misc-difference 92
FT                   /note= "MATURE N-TERMINUS"
XX
PN   WO8901940-A.
XX
PD   09-MAR-1989.
XX
PF   01-SEP-1988;   88WO-US002940.
XX
PR   04-SEP-1987;   87US-00094322.
PR   07-JAN-1988;   88US-00141649.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Fisher RA,  Gilbert W,  Sato VL,  Flavell RA,  Maraganore JM;
XX
DR   WPI; 1989-085519/11.
DR   N-PSDB; AAN90642.
XX
PT   DNA sequences coding for soluble T4-like polypeptide(s) - used in
PT   immuno:therapeutic and immunosuppressive compsns. and for preventing,
PT   treating or detecting AIDS.
XX
PS   Disclosure; Fig 3; 207pp; English.
XX
CC   The polypeptides encoded are useful in immunotherapeutic, prophylactic
CC   and diagnostic compsns. They can be used to purify HIV from a sample. The
CC   soluble T4 protein-based compsns. are useful in treating immunodeficient
CC   patients suffering from diseases caused by agents whose primary targets
CC   are T4+ lymphocytes. They can be used for preventing, treating or
CC   detecting AIDS, ARC and HIV infection. (Updated on 25-MAR-2003 to correct
CC   PR field.)
XX
SQ   Sequence 524 AA;

  Query Match             86.5%;  Score 2010;  DB 1;  Length 524;
  Best Local Similarity   88.8%;  Pred. No. 5.8e-129;
  Matches  403;  Conservative    2;  Mismatches   19;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         67 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 126

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        127 ILGNQGSFLTKGPSKLNDRADSRRSLRDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 186

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 246

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 306

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 QAERASSSKSWITSDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 366

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        367 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 426

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        427 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 462

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||     |    || :|||:  ||   | |
Db        463 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 491


RESULT 59
AAR07640
ID   AAR07640 standard; protein; 2458 AA.
XX
AC   AAR07640;
XX
DT   31-OCT-2002  (revised)
DT   20-DEC-1990  (first entry)
XX
DE   Deduced protein sequence of p170-2 comprising T4.
XX
KW   plasmid p170-2; soluble T4 protein; AIDS; ARC; HIV.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         400. .858
FT                   /label= T4 surface glycoprotein
XX
PN   WO9008198-A.
XX
PD   26-JUL-1990.
XX
PF   18-JAN-1989;   89US-00300096.
XX
PR   18-JAN-1989;   89US-00300096.
XX
PA   (HARD ) HARVARD COLLEGE.
```

```
XX
PI   Letvin NA;
XX
DR   WPI; 1990-254040/33.
DR   N-PSDB; AAQ05607.
XX
PT   Treating or preventing AIDS, ARC or HIV infection - by administering an
PT   immunologically effective amt. of soluble T4 protein.
XX
PS   Disclosure; Fig 1; 121pp; English.
XX
CC   Entire sequence from T4-encoding plasmid p170-2. It is almost identical
CC   to the sequence published by Maddon et al.(1985) with the exception of
CC   three codon changes. At T4 amino acid residue 3, (posn.403 of entire
CC   sequence) Lys is encoded in stead of Asn. At posn. 64, (posn.464) Arg
CC   replaces Trp and at posn. 231, (posn. 631) Ser replaces Phe. Soluble T4
CC   can be produced by truncating the CDS to remove the transmembrane and
CC   cytoplasmic domains. The soluble forms may be modified to increase their
CC   immunogenicity by addition of an adjuvant such as incomplete Freund's
CC   adjuvant. The T4 interferes with HIV/T4 interaction and elicits anti-
CC   soluble T4 antibody production. See also AAQ05608. (Updated on 31-OCT-
CC   2002 to add missing OS field.)
XX
SQ   Sequence 2458 AA;

  Query Match             86.5%;  Score 2010;  DB 2;  Length 2458;
  Best Local Similarity   88.8%;  Pred. No. 3.8e-128;
  Matches  403;  Conservative    2;  Mismatches   19;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        400 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 459

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        460 ILGNQGSFLTKGPSKLNDRADSRRSLRDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 519

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 579

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 639

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        640 QAERASSSKSWITSDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 699

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        700 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 759

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        760 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 795

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |   || :|||:  ||  | |
Db        796 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 824


RESULT 60
AAR04031
ID   AAR04031 standard; protein; 2458 AA.
XX
AC   AAR04031;
XX
DT   25-MAR-2003  (revised)
DT   31-OCT-2002  (revised)
DT   29-MAY-1990  (first entry)
XX
DE   Full length T4 encoded by plasmid p170-2.
XX
KW   Soluble T4; p170-2; anti-retroviral agent; AIDS; ARC; HIV; AZT.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 423
FT                   /note= "Trp of Madden et al replaced by Arg"
FT   Misc-difference 425
FT                   /note= "Asp of Madden et al replaced by Lys"
FT   Misc-difference 653
FT                   /note= "Phe of Madden et al replaced by Ser"
XX
PN   WO8911860-A.
XX
PD   14-DEC-1989.
XX
PF   08-JUN-1989;   89WO-US002453.
XX
PR   10-JUN-1988;   88US-00204645.
PR   20-APR-1989;   89US-00341080.
XX
PA   (BIOJ ) BIOGEN NV INC.
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (BIOJ ) BIOGEN INC.
PA   (BIOJ ) BIOGEN INC.
XX
PI   Fisher RA,  Schooley RT,  Hirsch MS,  Johnson VA,  Walker BD;
XX
DR   WPI; 1990-007302/01.
DR   N-PSDB; AAQ03005.
XX
PT   Combinations of soluble T4 protein and anti-retro:viral agent - having
PT   synergistic activity in treatment and prevention of AIDS, arc and HIV
PT   infection.
XX
PS   Disclosure; Fig 1; 100pp; English.
XX
CC   The sequence differs from that determined by PJ Madden et al., [Cell, 42
CC   pp. 93-104 (1985)] in three places due to three nucleotide substitutions.
CC   The Asp reported at position 3 by Madden et al. was the result of a
CC   sequencing error [DR Littman et al. Cell, 55, p.541 (1988)]. X = stop
CC   codon. The sequence was deduced from the cDNA insert of p170-2. Soluble
CC   T4 constructs may be produced by truncating this sequence to give
CC   fragments from position 400 to 799, removing the transmembrane and
CC   intracytoplasmic domains whilst retaining the extracellular region
CC   responsible for HIV binding. The sol. T4 is combined with an anti-viral
CC   agent such as AZT. See also AAQ03006. (Updated on 31-OCT-2002 to add
CC   missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 2458 AA;

  Query Match             86.5%;  Score 2010;  DB 2;  Length 2458;
  Best Local Similarity   88.8%;  Pred. No. 3.8e-128;
  Matches  403;  Conservative    2;  Mismatches   19;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        400 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 459

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        460 ILGNQGSFLTKGPSKLNDRADSRRSLRDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 519

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 579

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 639

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
```

```
                ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 QAERASSSKSWITSDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 699

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        700 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 759

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        760 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 795

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||    |   || :|||:  ||  | |
Db        796 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 824


RESULT 61
AAR11285
ID   AAR11285 standard; protein; 458 AA.
XX
AC   AAR11285;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   29-APR-1991  (first entry)
XX
DE   gp120 binding protein.
XX
KW   Human; CD4; AIDS; HIV1; SIV; gp120.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 59
FT                   /label= Thr or Ile
FT   Misc-difference 93
FT                   /label= Thr or Pro
XX
PN   EP414178-A.
XX
PD   27-FEB-1991.
XX
PF   18-AUG-1990;   90EP-00115877.
XX
PR   23-AUG-1989;   89US-00397782.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Camerini D;
XX
DR   WPI; 1991-059419/09.
DR   N-PSDB; AAQ10887.
XX
PT   New non-human primate and human CD4 or gp120 molecules - used to treat
PT   HIV or SIV and immunoglobulin and gp120 binding molecules from new fusion
PT   proteins.
XX
PS   Claim 17; Page 57; 87pp; English.
XX
CC   The fragment from residues 1-134 is also independently claimed. The sub-
CC   fragment (and the complete polypeptide) can bind to HIV gp120. See also
CC   AAQ10885-6, AAQ10888. (Updated on 09-JAN-2003 to add missing OS field.)
CC   (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 458 AA;

  Query Match             86.4%;  Score 2008;  DB 2;  Length 458;
  Best Local Similarity   88.3%;  Pred. No. 6.7e-129;
  Matches  401;  Conservative    3;  Mismatches   20;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||| |
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQXK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||||||||||||||| |||||||||||||||| ||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFXLIIKNLKIEDSDTYICEVGDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 EAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||    |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 62
AAR10988
ID   AAR10988 standard; protein; 458 AA.
XX
AC   AAR10988;
XX
DT   25-MAR-2003  (revised)
DT   29-APR-1991  (first entry)
XX
DE   Chimpanzee CD4 protein.
XX
KW   Chimpanzee; CD4; AIDS; HIV1; SIV.
XX
OS   Pan troglodytes.
XX
FH   Key             Location/Qualifiers
FT   Protein         26
FT                   /label= mature CD4
XX
PN   EP414178-A.
XX
PD   27-FEB-1991.
XX
PF   18-AUG-1990;   90EP-00115877.
XX
PR   23-AUG-1989;   89US-00397782.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Camerini D;
XX
DR   WPI; 1991-059419/09.
DR   N-PSDB; AAQ10886.
XX
PT   New non-human primate and human CD4 or gp120 molecules - used to treat
PT   HIV or SIV and immunoglobulin and gp120 binding molecules from new fusion
PT   proteins.
XX
PS   Claim 4; Page 45; 87pp; English.
XX
CC   The CD4 protein or HIV gp120-binding fragments of it are used to detect
CC   and treat HIV and SIV infection. Animals which can be treated include
CC   humans, baboons, orang-utans, chimpanzees, gorillas and rhesus monkeys.
```

```
CC   The chimpanzee CD4 is 99 per cent homologous to its human counterpart,
CC   possessing 5 amino acid substitutions in the 433 residue predicted mature
CC   polypeptide. See also AAQ10885, AAQ10887-8. (Updated on 25-MAR-2003 to
CC   correct PA field.)
XX
SQ   Sequence 458 AA;

  Query Match             86.4%;  Score 2008;  DB 2;  Length 458;
  Best Local Similarity   88.3%;  Pred. No. 6.7e-129;
  Matches  401;  Conservative    2;  Mismatches   21;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||| |
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||| |||||||||||| |||||||||||||||||| ||||||||
Db         61 ILGNQGSFLTKGPSKLNDRVDSRRSLWDQGNFTLIIKNLKIEDSDTYICEVGDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP----------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |    || :|||:  ||   | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 63
AAR04910
ID   AAR04910 standard; protein; 458 AA.
XX
AC   AAR04910;
XX
DT   31-OCT-2002  (revised)
DT   02-OCT-1990  (first entry)
XX
DE   T4 protein as encoded by p170.2.
XX
KW   T4 protein; immunotoxin; Pseudomona endotoxin A; AIDS; HIV; ARC;
KW   angiogenin; fusion protein.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .26
FT                   /label= signal peptide
FT   Protein         27. .458
FT                   /label= T4 protein
FT   Misc-difference replace(87,W)
FT                   /note= "differs from Madden et al"
FT   Misc-difference replace(254,F)
FT                   /note= "differs from Madden et al"
XX
PN   WO9004414-A.
XX
PD   03-MAY-1990.
XX
PF   18-OCT-1988;   88US-00259355.
XX
PR   18-OCT-1988;   88US-00259355.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Meade HM,  Lobb RR,  Gates LL,  Winkler G;
XX
DR   WPI; 1990-163876/21.
DR   N-PSDB; AAQ04555.
XX
PT   New immunotoxin contg. soluble T4 protein components and toxin - esp.
PT   Pseudomonas endotoxin A, for treating or controlling AIDS and related
PT   conditions, and new DNA sequences.
XX
PS   Disclosure; Page ?; -pp; English.
XX
CC   The T4 protein encoded by p170.2 is almost identical to that reported by
CC   P.J. Maddon et al [Cell, 42, pp 93-104 (1985)]. The Maddon sequence was
CC   revised in 1988 to correct a DNA sequencing error at AA 3 (corrected from
CC   Asp to Lys; see M12807 in GenBank). The DNA may be truncated (to remove
CC   transmembrane and intracellular regions) and/or modified by SDM, pref. so
CC   the prod. extends from AAs 3-183 of the mature protein. This DNA can then
CC   be ligated to a toxin DNA esp. angiogenin, or a fragment of Pseudomonas
CC   exotoxin A contg. the translocation and ADP-ribosylation domains. The
CC   hybrid DNA can then be inserted into an expression vector and used to
CC   produce recombinant fusion protein which is useful for preventing or
CC   treating AIDS, ARC, and HIV infections. The T4 protein is an HIV receptor
CC   which binds to virus or to infected cells carrying the gp120/160 marker
CC   antigen, so provides v. specific targetting with minimal damage to non-
CC   target cells. (Updated on 31-OCT-2002 to add missing OS field.)
XX
SQ   Sequence 458 AA;

  Query Match             86.1%;  Score 2002;  DB 2;  Length 458;
  Best Local Similarity   88.3%;  Pred. No. 1.7e-128;
  Matches  401;  Conservative    4;  Mismatches   19;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLRDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              |||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTASSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITSDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||:|||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIRVLPTWSTPVQP----------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |    || :|||:  ||   | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 64
```

```
AAB07768
ID   AAB07768 standard; protein; 394 AA.
XX
AC   AAB07768;
XX
DT   07-NOV-2000  (first entry)
XX
DE   The soluble extracellular domain of the T4 glycoprotein.
XX
KW   Human; T4 glycoprotein; human immunodeficiency virus; HIV;
KW   envelope glycoprotein; AIDS; virus binding.
XX
OS   Homo sapiens.
XX
PN   US6093539-A.
XX
PD   25-JUL-2000.
XX
PF   06-JUN-1995;   95US-00466368.
XX
PR   21-AUG-1986;   86US-00898587.
PR   11-JUN-1991;   91US-00713564.
PR   06-JUL-1992;   92US-00909021.
PR   12-DEC-1994;   94US-00354452.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Maddon PJ,  Chess L,  Axel R,  Weiss R,  Mcdougal JS,  Littman DR;
XX
DR   WPI; 2000-505203/45.
DR   N-PSDB; AAA59351.
XX
PT   New isolated nucleic acid encoding a human T cell surface protein and the
PT   soluble surface T4 glycoprotein that it encodes, useful as prophylaxis
PT   for treating a subject infected with human acquired immune deficiency
PT   syndrome virus.
XX
PS   Disclosure; Col 11-14; 69pp; English.
XX
CC   The present sequence represents an aqueous-soluble polypeptide comprising
CC   a portion of a human T4 glycoprotein. The portion specifically forms a
CC   complex with a human immunodeficiency virus (HIV) envelope glycoprotein.
CC   The DNA is useful for producing the soluble surface T4 glycoprotein. The
CC   soluble surface T4 glycoprotein is useful as a therapeutic agent, i.e. as
CC   prophylaxis for treating a subject infected with an HIV virus. Thus, the
CC   soluble T4 glycoprotein is useful for treating human AIDS. The soluble T4
CC   glycoprotein is also useful in diagnostic or screening assays, e.g. for
CC   screening inhibitors of virus binding, or for detecting and quantitating
CC   T4, T4+ cells and antibodies to T4, which are of diagnostic value for
CC   AIDS
XX
SQ   Sequence 394 AA;

  Query Match             86.1%;  Score 2001;  DB 3;  Length 394;
  Best Local Similarity   98.7%;  Pred. No. 1.7e-128;
  Matches  389;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||| |||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db         61 ILGNQGSSLTKGPSKLNDRADSRRSLWDQGNFPLIIRNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVDFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394


RESULT 65
AAR20151
ID   AAR20151 standard; protein; 399 AA.
XX
AC   AAR20151;
XX
DT   25-MAR-2003  (revised)
DT   31-MAR-1992  (first entry)
XX
DE   Chimpanzee sol. CD4 encoded by pSQ200.
XX
KW   Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
KW   acquired immune deficiency syndrome; AIDS related complex;
KW   T helper lymphocytes.
XX
OS   Pan troglodytes.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .25
FT                   /label= signal_sequence
XX
PN   WO9118618-A.
XX
PD   12-DEC-1991.
XX
PF   25-MAY-1990;   90US-00529186.
XX
PR   25-MAY-1990;   90US-00529186.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Fisher RA,  Hession C,  Burkly LC;
XX
DR   WPI; 1992-007200/01.
DR   N-PSDB; AAQ20326.
XX
PT   New immuno-therapeutic human CD4 variants and derivs. - elicit AB
PT   production to HIV gp.120, useful in treating, preventing and diagnosing
PT   AIDS, ARC and HIV infections.
XX
PS   Claim 15; Fig 21; 179pp; English.
XX
CC   The sequence was deduced from the DNA sequence of clone pSQ200 which was
CC   obtd. by cloning using a reverse transcriptase/PCR amplification
CC   procedure with a non-specific oligo dT to prime first strand synthesis.
CC   The PCR fragments were ligated into sequencing plasmid pNN08 (see
CC   AAQ20326) The DNA sequence can be used to express recombinant soluble CD4
CC   and analogues for use in diagnosis and treatment of diseases caused by
CC   infective agents whose primary targets are T4+ lymphocytes. See also
CC   AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA
CC   field.)
XX
SQ   Sequence 399 AA;

  Query Match             86.1%;  Score 2001;  DB 2;  Length 399;
  Best Local Similarity   98.0%;  Pred. No. 1.7e-128;
  Matches  390;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||| |
```

```
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQTK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||| |||||||||||| |||||||||||||||||| ||||||||
Db         61 ILGNQGSFLTKGPSKLNDRVDSRRSLWDQGNFTLIIKNLKIEDSDTYICEVGDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRA 398
              |||||||||||||||||||||||||||||||| | ||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMA 398
```

RESULT 66
AAR20150
ID AAR20150 standard; protein; 400 AA.
XX
AC AAR20150;
XX
DT 25-MAR-2003 (revised)
DT 31-MAR-1992 (first entry)
XX
DE Chimpanzee sol. CD4 sequence from pSQ205.
XX
KW Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
KW acquired immune deficiency syndrome; AIDS related complex;
KW T helper lymphocytes.
XX
OS Pan troglodytes.
XX
FH Key Location/Qualifiers
FT Peptide 1. .25
FT /label= signal_sequence
XX
PN WO9118618-A.
XX
PD 12-DEC-1991.
XX
PF 25-MAY-1990; 90US-00529186.
XX
PR 25-MAY-1990; 90US-00529186.
XX
PA (BIOJ) BIOGEN INC.
XX
PI Fisher RA, Hession C, Burkly LC;
XX
DR WPI; 1992-007200/01.
DR N-PSDB; AAQ20325.
XX
PT New immuno-therapeutic human CD4 variants and derivs. - elicit AB
PT production to HIV gp.120, useful in treating, preventing and diagnosing
PT AIDS, ARC and HIV infections.
XX
PS Claim 15; Fig 20; 179pp; English.
XX
CC The sequence was deduced from the DNA sequence of clone pSQ205 which was
CC obtd. by cloning using a reverse transcriptase/PCR amplification
CC procedure. The DNA sequence can be used to express recombinant soluble
CC CD4 and analogues for use in diagnosis and treatment of diseases caused
CC by infective agents whose primary targets are T4+ lymphocytes. See also
CC AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA
CC field.)
XX
SQ Sequence 400 AA;

Query Match 85.8%; Score 1996; DB 2; Length 400;
Best Local Similarity 97.7%; Pred. No. 3.8e-128;
Matches 389; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||| |||||||||||||||||||||||||||||:| |
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSSQTK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||| |||||||||||| |||||||||||||||||| ||||||||
Db         61 ILGNQGSFLTKGPSKLNDRVDSRRSLWDQGNFTLIIKNLKIEDSDTYICEVGDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRA 398
              |||||||||||||||||||||||||||||||| | ||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMA 398
```

RESULT 67
AAP93010
ID AAP93010 standard; protein; 399 AA.
XX
AC AAP93010;
XX
DT 25-MAR-2003 (revised)
DT 03-AUG-1992 (first entry)
XX
DE Genetic construct which encodes CD4 linked to human IgM at the Mst2 site
DE upstream of the CH1 region (fusion protein CD4Mmu).
XX
KW Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
KW diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
XX
OS Homo sapiens.
XX
PN EP325262-A.
XX
PD 26-JUL-1989.
XX
PF 20-JAN-1989; 89EP-00100913.
XX
PR 22-JAN-1988; 88US-00147351.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 1989-214472/30.
DR N-PSDB; AAN90358.
XX
PT Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PT infections or detecting HIV or SIV in sample.

```
XX
PS   Example; Table 3, Page 34-40; 68pp; English.
XX
CC   The fusion protein genes of the invention pref. comprises cDNA sequences
CC   which encode CD4 or a fragment which binds gp120 ligated to an expression
CC   plasmid which encodes an antibody in which the variable region of the
CC   gene has been deleted (see Wo87-02671). The CD4 portion of the fusion
CC   protein may comprise the complete CD4 sequence, the 370 AA extracellular
CC   region and the membrane spanning domain, or the extracellular region. The
CC   Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are
CC   specifically claimed: fusion proteins CD4Hlambda1, CD4Mmu, CD4Pmu,
CC   CD4Elambda1, and CD4Mmu (No. 67608), pCD4Plambda (No. 67609) and
CC   pCD4Elambda1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 399 AA;

  Query Match             85.6%;  Score 1990;  DB 1;  Length 399;
  Best Local Similarity   98.5%;  Pred. No. 9.6e-128;
  Matches  389;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| |||||||||||||||||||||||| |||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHCKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||| ||| |||| ||:||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTAKSDTLLLQGLSLSLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 68
AAB19511
ID   AAB19511 standard; protein; 616 AA.
XX
AC   AAB19511;
XX
DT   09-JAN-2001  (first entry)
XX
DE   CD4-IgG1 fusion protein CH4Bgamma1.
XX
KW   CD4; IgG1; human; CD4Bgamma1; fusion protein; immunoglobulin; HIV; SIV;
KW   gp120; therapy; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         1. .395
FT                   /note= "CD4 extracellular region"
FT   Protein         400. .616
FT                   /note= "IgG1 heavy chain"
XX
PN   US6117656-A.
XX
PD   12-SEP-2000.
XX
PF   07-JUN-1995;   95US-00479353.
XX
PR   22-JAN-1988;   88US-00147351.
PR   23-JAN-1989;   89US-00299596.
PR   09-JUN-1992;   92US-00896781.
PR   12-APR-1993;   93US-00057952.
PR   04-FEB-1994;   94US-00191708.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B;
XX
DR   WPI; 2000-586558/55.
DR   N-PSDB; AAA50664.
XX
PT   CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
PT   SIV.
XX
PS   Example 1; Col 59-70; 39pp; English.
XX
CC   The present sequence is that of fusion protein CD4Bgamma1 comprising the
CC   extracellular portion of CD4, which binds to HIV gp120, linked at its C-
CC   terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
CC   encoding CD4 was linked to IgG1 DNA at the Ban1 site downstream of the
CC   hinge region (see AAA50664). Fusion protein CD4Bgamma1 and a nucleic acid
CC   encoding it are claimed. Also claimed are a vector comprising the nucleic
CC   acid, and a method of producing the fusion protein in secreted form using
CC   a transformed host cell. The fusion protein may further comprise a
CC   therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
CC   can be administered to an animal (including humans) for treatment of HIV
CC   or SIV infection, and can also be used in assays for HIV or SIV, imaging
CC   and tissue stains. IgG1 fusion proteins such as CD4Bgamma1 provide both
CC   complement-mediated and cell-mediated immunity
XX
SQ   Sequence 616 AA;

  Query Match             84.9%;  Score 1974;  DB 3;  Length 616;
  Best Local Similarity   94.9%;  Pred. No. 2e-126;
  Matches  392;  Conservative    1;  Mismatches   18;  Indels    2;  Gaps    1;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||| |||||||||||||||||||| |||||||||| ||||  |||||||||||||
Db        121 LVFCLTANSDTHLLQGQSLTLTLESPPCSSPSVQCRSPRCKNIQCCKTLSVSQLELQDSC 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTCSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PRASALPAPPTGSALP 411
              |||||||||||||||||||||||||||||||||||  | |  |  |:  |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPEAPELLGGPSVFLFP 413


RESULT 69
AAR20154
ID   AAR20154 standard; protein; 400 AA.
```

```
XX
AC   AAR20154;
XX
DT   24-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   31-MAR-1992  (first entry)
XX
DE   Sol. rhesus-human chimeric CD4 encoded by pDG100.
XX
KW   Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
KW   acquired immune deficiency syndrome; AIDS related complex; monkey;
KW   T helper lymphocytes.
XX
OS   Macaca mulatta; (Rhesus).
OS   Homo sapiens.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .131
FT                   /note= "rhesus CD4 encoded by pSQ146 (AAQ20328)"
FT   Peptide         1. .25
FT                   /label= signal_sequence
FT   Region          132. .400
FT                   /note= "human CD4 encoded by pBG391 (US8802940)"
XX
PN   WO9118618-A.
XX
PD   12-DEC-1991.
XX
PF   25-MAY-1990;   90US-00529186.
XX
PR   25-MAY-1990;   90US-00529186.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Fisher RA,  Hession C,  Burkly LC;
XX
DR   WPI; 1992-007200/01.
DR   N-PSDB; AAQ20329.
XX
PT   New immuno-therapeutic human CD4 variants and derivs. - elicit AB
PT   production to HIV gp.120, useful in treating, preventing and diagnosing
PT   AIDS, ARC and HIV infections.
XX
PS   Claim 14; Fig 13; 179pp; English.
XX
CC   The sequence is encoded by pDG100, chimeric plasmid prepd. from human and
CC   rhesus derived CD4 DNA. The plasmid can be used to express recombinant
CC   sol. chimeric CD4 for use in diagnosis and treatment of diseases caused
CC   by infective agents whose primary targets are T4+ lymphocytes. See also
CC   AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA
CC   field.) (Updated on 24-OCT-2003 to standardise OS field)
XX
SQ   Sequence 400 AA;

  Query Match             84.1%;  Score 1956;  DB 2;  Length 400;
  Best Local Similarity   95.2%;  Pred. No. 2e-125;
  Matches  379;  Conservative    9;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||:|||||||||||||||| ||| |||||||||||||| |||||: ||||||||||
Db          1 MNRGIPFRHLLLVLQLALLPAVTQGKKVVLGKKGDTVELTCNASQKKNTQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||| |||||||||||:||||||:|||||| | :||||||||||||||::||||:|
Db         61 ILGIQGSFLTKGPSKLSDRADSRKSLWDQGCFSMIIKNLKIEDSDTYICEVENKKEEVEL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLEGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRA 398
              |||||||||||||||||||||||||||||| | |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQPMA 398


RESULT 70
AAR15149
ID   AAR15149 standard; protein; 458 AA.
XX
AC   AAR15149;
XX
DT   25-MAR-2003  (revised)
DT   24-FEB-1992  (first entry)
XX
DE   CD4 coordinate system.
XX
KW   Gelsolin; fusion protein; diagnosis; AIDS.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1. .25
FT                   /label= hydrophobic/secretory_signal
FT   Domain          26. .132
FT                   /label= first_Ig-related_domain
FT                   /note= "extracellular"
FT   Disulfide-bond  41. .109
FT   Domain          133. .202
FT                   /label= second_Ig-related_domain
FT                   /note= "extracellular"
FT   Disulfide-bond  155. .184
FT   Domain          203. .318
FT                   /label= third_IG-related_domain
FT                   /note= "extracellular"
FT   Domain          319. .395
FT                   /label= fourth_Ig-related_domain
FT                   /note= "extracellular"
FT   Disulfide-bond  328. .370
FT   Domain          396. .416
FT                   /label= hydrophobic/transmembrane_sequence
FT   Domain          417. .458
FT                   /label= very_hydrophilic/intracytoplasmic
XX
PN   WO9117170-A.
XX
PD   14-NOV-1991.
XX
PF   04-MAY-1990;   90US-00520368.
XX
PR   04-MAY-1990;   90US-00520368.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Pepinsky RB,  Rosa MD,  Stossel TP;
XX
DR   WPI; 1991-353711/48.
DR   N-PSDB; AAQ14931.
XX
PT   New multi-meric and hetero-multi-meric gelsolin fusion constructs - used
PT   to treat and diagnose AIDS, ARC and HIV infection.
XX
PS   Disclosure; Fig 3A-3D; 131pp; English.
XX
```

```
CC   The CD4 polypeptides useful in the constructs include all CD4
CC   polypeptides which bind to or otherwise inhibit gp120 and gp160. These
CC   include fragments lacking the transmembrane domain. In particular it is
CC   CD4 1-111; 1-Cys111; 1-Cys180; 1-181, 1-183; 1-187; 1-345 or 1-375 (from
CC   mature protein). See also AAQ14931-35 and AAR15151. (Updated on 25-MAR-
CC   2003 to correct PA field.)
XX
SQ   Sequence 458 AA;

  Query Match             83.2%;  Score 1934;  DB 2;  Length 458;
  Best Local Similarity   85.2%;  Pred. No. 7.6e-124;
  Matches  387;  Conservative    9;  Mismatches   28;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              |||||||  :  :::|: ::||||   ||||||||||||||||||||||||||||||||
Db        121 LVFGLTAKCERDVVEGERVSLTLERGGGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITSDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||| |||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLRTWSTPVQP------------------------ 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                      |    |   || :|||:  ||  | |
Db        397 -----MGLIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 71
AAP90833
ID   AAP90833 standard; protein; 384 AA.
XX
AC   AAP90833;
XX
DT   25-MAR-2003  (revised)
DT   01-AUG-1990  (first entry)
XX
DE   Amino acid sequence of a soluble T4-like (sT4) polypeptide encoded by a
DE   portion of clone p199-7 (PL mutet.rsT4).
XX
KW   HIV; immunotherapeutic; prophylactic; soluble T4-like polypeptide;
KW   diagnostic; p199-7 (PL mutet.rsT4).
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 10
FT                   /note= "initiating Met"
XX
PN   WO8901940-A.
XX
PD   09-MAR-1989.
XX
PF   01-SEP-1988;   88WO-US002940.
XX
PR   04-SEP-1987;   87US-00094322.
PR   07-JAN-1988;   88US-00141649.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Fisher RA,  Gilbert W,  Sato VL,  Flavell RA,  Maraganore JM;
XX
DR   WPI; 1989-085519/11.
DR   N-PSDB; AAN90643.
XX
PT   DNA sequences coding for soluble T4-like polypeptide(s) - used in
PT   immuno:therapeutic and immunosuppressive compsns. and for preventing,
PT   treating or detecting AIDS.
XX
PS   Disclosure; Page ?; 207pp; English.
XX
CC   It is the protein sequence encoded by the rsT4 sequence. It is claimed in
CC   the patent. It is useful in immunotherapeutic, prophylactic and
CC   diagnostic compsns. It can be used to purify HIV from a sample. (Updated
CC   on 25-MAR-2003 to correct PR field.)
XX
SQ   Sequence 384 AA;

  Query Match             82.4%;  Score 1915;  DB 1;  Length 384;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-122;
  Matches  371;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 70

Qy         84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 143
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 130

Qy        144 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 190

Qy        204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 250

Qy        264 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 323
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 310

Qy        324 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESN 383
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        311 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESN 370

Qy        384 IKVLPTWSTPVHP 396
              |||||||||||| |
Db        371 IKVLPTWSTPVQP 383


RESULT 72
AAR41041
ID   AAR41041 standard; protein; 942 AA.
XX
AC   AAR41041;
XX
DT   24-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   22-MAR-1994  (first entry)
XX
DE   CD4-GBP130 fusion protein.
XX
KW   Merozoite; Glycophorin Binding Protein 130; malaria; HIV; env;
KW   human immunodeficiency virus; envelope glycoprotein; hybrid protein;
KW   red blood cell; erythrocyte; AIDS.
XX
OS   Homo sapiens.
```

```
OS   Plasmodium falciparum.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .371
FT                   /note= "residues 1-371 of CD4"
FT   Region          372. .942
FT                   /note= "residues 201-774 of GBP130"
XX
PN   WO9318160-A1.
XX
PD   16-SEP-1993.
XX
PF   10-MAR-1993;   93WO-GB000505.
XX
PR   11-MAR-1992;   92GB-00005276.
PR   08-JUL-1992;   92GB-00014481.
PR   24-JUL-1992;   92GB-00015829.
PR   16-SEP-1992;   92GB-00019562.
PR   03-MAR-1993;   93GB-00004311.
XX
PA   (PREN/) PRENDERGAST K F.
XX
PI   Prendergast KF;
XX
DR   WPI; 1993-303474/38.
XX
PT   Anti-viral fusion peptide(s) - comprise viral-binding component and
PT   malaria merozoite red cell binding component, for treating e.g. HIV, and
PT   hepatitis.
XX
PS   Claim 2; Page 35-37; 69pp; English.
XX
CC   The hybrid protein NH2-CD4(1-371)-GBP130(201-774)-COOH is a specifically
CC   claimed example of a fusion protein of the invention; it comprises at
CC   least part of the CD4 molecule fused to a peptide from a malarial
CC   parasite merozoite protein with affinity for red blood cells. The fusion
CC   protein can bind free HIV in the blood to red blood cells and
CC   consequently reduce viral titre, prevent transmission of the virus and
CC   improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct
CC   PN field.) (Updated on 24-OCT-2003 to standardise OS field)
XX
SQ   Sequence 942 AA;

  Query Match             82.2%;  Score 1911;  DB 2;  Length 942;
  Best Local Similarity   98.7%;  Pred. No. 6.9e-122;
  Matches  371;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 60

Qy         84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 120

Qy        144 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 180

Qy        204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 240

Qy        264 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 300

Qy        324 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESN 383
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        301 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESN 360

Qy        384 IKVLPTWSTPVHPRAS 399
              ||||||||||| : |
Db        361 IKVLPTWSTPVSQKPS 376


RESULT 73
AAR41043
ID   AAR41043 standard; protein; 1786 AA.
XX
AC   AAR41043;
XX
DT   24-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   22-MAR-1994  (first entry)
XX
DE   CD4-EBA175 fusion protein.
XX
KW   Merozoite; Erythrocyte Binding Antigen 175; malaria; HIV; env;
KW   human immunodeficiency virus; envelope glycoprotein; hybrid protein;
KW   red blood cell; erythrocyte; AIDS; molecular machine.
XX
OS   Homo sapiens.
OS   Plasmodium falciparum.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .371
FT                   /note= "residues 1-371 of CD4"
FT   Region          372. .1786
FT                   /note= "residues 20-1435 of EBA-175"
XX
PN   WO9318160-A1.
XX
PD   16-SEP-1993.
XX
PF   10-MAR-1993;   93WO-GB000505.
XX
PR   11-MAR-1992;   92GB-00005276.
PR   08-JUL-1992;   92GB-00014481.
PR   24-JUL-1992;   92GB-00015829.
PR   16-SEP-1992;   92GB-00019562.
PR   03-MAR-1993;   93GB-00004311.
XX
PA   (PREN/) PRENDERGAST K F.
XX
PI   Prendergast KF;
XX
DR   WPI; 1993-303474/38.
XX
PT   Anti-viral fusion peptide(s) - comprise viral-binding component and
PT   malaria merozoite red cell binding component, for treating e.g. HIV, and
PT   hepatitis.
XX
PS   Claim 9; Page 44-47; 69pp; English.
XX
CC   The hybrid protein NH2-CD4(1-371)-EBA175(20-1435)-COOH is a specifically
CC   claimed example of a fusion protein of the invention; it comprises at
CC   least part of the CD4 molecule fused to a peptide from a malarial
CC   parasite merozoite protein with affinity for red blood cells. The fusion
CC   protein can bind free HIV in the blood to red blood cells and
CC   consequently reduce viral titre, prevent transmission of the virus and
CC   improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct
CC   PN field.) (Updated on 24-OCT-2003 to standardise OS field)
XX
SQ   Sequence 1786 AA;

  Query Match             82.2%;  Score 1911;  DB 2;  Length 1786;
  Best Local Similarity   99.2%;  Pred. No. 1.5e-121;
  Matches  371;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 60
```

```
Qy          84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 143
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 120

Qy         144 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 180

Qy         204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 240

Qy         264 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 323
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 300

Qy         324 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESN 383
               |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         301 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESN 360

Qy         384 IKVLPTWSTPVHPR 397
               |||||||||||  |
Db         361 IKVLPTWSTPVKAR 374
```

```
RESULT 74
AAW41376
ID   AAW41376 standard; peptide; 433 AA.
XX
AC   AAW41376;
XX
DT   28-MAY-1998  (first entry)
XX
DE   Human CD4.
XX
KW   Antibody; CD4; passive immunity; HIV type 1; HIV type 2; HIV infection;
KW   simian immunodeficiency virus; SIV; AIDS; therapy; HIV gp120.
XX
OS   Homo sapiens.
XX
PN   WO9746697-A2.
XX
PD   11-DEC-1997.
XX
PF   03-JUN-1997;   97WO-US009449.
XX
PR   03-JUN-1996;   96US-00657149.
PR   28-FEB-1997;   97US-00808374.
PR   02-JUN-1997;   97US-00867149.
XX
PA   (UNBI-) UNITED BIOMEDICAL INC.
XX
PI   Wang CY;
XX
DR   WPI; 1998-042204/04.
XX
PT   Antibody against complex of CD4 and chemokine receptor domain - useful
PT   for prevention and treatment of human immunodeficiency virus infection.
XX
PS   Disclosure; Page 122-123; 140pp; English.
XX
CC   This sequence represents human CD4, and was used to isolate the antibody
CC   (Ab) of the invention. The Ab, preferably a M2 or B13 Ab, has the
CC   following characteristics: (a) binds rsCD4; (b) binds CD4 expressing
CC   cells in an immunofluorescence assay, where the binding pattern is in the
CC   shape of ''caps'' when examined with a high resolution fluorescence
CC   microscope; (c) blocks the binding of HIV glycoprotein 120 (gp120) to CD4
CC   expressing cells; (d) binds CD4 expressing cells previously bound with
CC   gp120; and (e) neutralises HIV primary isolates in an in vitro
CC   microplaque assay at a concentration of less than 10 mu g/ml, preferably
CC   at a concentration in the range of 0.01-10 mu g/ml for 50% neutralisation
CC   and 0.1-35 mu g/ml for 90% neutralisation. The Ab can be used to provide
CC   passive immunity to HIV in a mammal, when administered parenterally,
CC   specifically all clades of HIV type 1, and from diverse primary isolates
CC   of HIV type 2 and simian immunodeficiency virus (SIV). The Ab is
CC   prophylactic and therapeutic for HIV infection and all stages of AIDS
CC   because it prevents replicative infection of host cells both before and
CC   after HIV gp120 has bound to the host cell antigen complex comprising CD4
CC   on the surface of CD4 positive lymphocytes, thus it is capable of
CC   preventing HIV infection and retarding the spread of the virus to
CC   uninfected cells. It is also uniquely useful because it inhibits
CC   infection following binding of HIV to CD4 expressing cells
XX
SQ   Sequence 433 AA;
```

```
  Query Match             82.1%;  Score 1909;  DB 2;  Length 433;
  Best Local Similarity   88.6%;  Pred. No. 3.6e-122;
  Matches  380;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          26 NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS 85
               |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db           1 NKVVLGKKGDTVELTCTASQKKSIQFHWKNWNQIKILGNQGSFLTKGPSKLNDRADSRRS 60

Qy          86 LWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLES 145
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLES 120

Qy         146 PPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQ 205
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 PPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQ 180

Qy         206 KASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKR 265
               |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         181 KASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIIFDLKNKEVSVKR 240

Qy         266 VTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKN 325
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 VTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKN 300

Qy         326 LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIK 385
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIK 360

Qy         386 VLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGV 445
               ||||||||| |                                 ||     |   || :|||:
Db         361 VLPTWSTPVQP-----------------------------MALIVLGGVAGLLLFIGLGI 391

Qy         446 -ACVLARTR 453
                || | |
Db         392 FFCVRCRHR 400
```

```
RESULT 75
AAR41042
ID   AAR41042 standard; protein; 729 AA.
XX
AC   AAR41042;
XX
DT   24-OCT-2003  (revised)
DT   25-MAR-2003  (revised)
DT   22-MAR-1994  (first entry)
XX
DE   CD4-GBPH fusion protein.
XX
KW   Merozoite; Glycophorin Binding Protein homologue; malaria; HIV; env;
KW   human immunodeficiency virus; envelope glycoprotein; hybrid protein;
KW   red blood cell; erythrocyte; AIDS.
XX
OS   Homo sapiens.
OS   Plasmodium falciparum.
OS   Chimeric.
XX
FH   Key             Location/Qualifiers
FT   Region          1. .371
FT                   /note= "residues 1-371 of CD4"
FT   Region          372. .729
```

```
FT                   /note= "residues 70-427 of GBPH"
XX
PN   WO9318160-A1.
XX
PD   16-SEP-1993.
XX
PF   10-MAR-1993;   93WO-GB000505.
XX
PR   11-MAR-1992;   92GB-00005276.
PR   08-JUL-1992;   92GB-00014481.
PR   24-JUL-1992;   92GB-00015829.
PR   16-SEP-1992;   92GB-00019562.
PR   03-MAR-1993;   93GB-00004311.
XX
PA   (PREN/) PRENDERGAST K F.
XX
PI   Prendergast KF;
XX
DR   WPI; 1993-303474/38.
XX
PT   Anti-viral fusion peptide(s) - comprise viral-binding component and
PT   malaria merozoite red cell binding component, for treating e.g. HIV, and
PT   hepatitis.
XX
PS   Claim 7; Page 40-41; 69pp; English.
XX
CC   The hybrid protein NH2-CD4(1-371)-GBPH(70-427)-COOH is a specifically
CC   claimed example of a fusion protein of the invention; it comprises at
CC   least part of the CD4 molecule fused to a peptide from a malarial
CC   parasite merozoite protein with affinity for red blood cells. The fusion
CC   protein can bind free HIV in the blood to red blood cells and
CC   consequently reduce viral titre, prevent transmission of the virus and
CC   improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct
CC   PN field.) (Updated on 24-OCT-2003 to standardise OS field)
XX
SQ   Sequence 729 AA;

  Query Match             82.1%;  Score 1908;  DB 2;  Length 729;
  Best Local Similarity   99.7%;  Pred. No. 8e-122;
  Matches  370;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 60

Qy         84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 120

Qy        144 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 180

Qy        204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 240

Qy        264 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 300

Qy        324 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESN 383
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        301 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESN 360

Qy        384 IKVLPTWSTPV 394
              ||||||||||| 
Db        361 IKVLPTWSTPV 371


RESULT 76
AAR07721
ID   AAR07721 standard; protein; 375 AA.
XX
AC   AAR07721;
XX
DT   25-MAR-2003  (revised)
DT   18-FEB-1991  (first entry)
XX
DE   Recombinant soluble (rs) T4.
XX
KW   Metal-binding site.
XX
OS   Homo sapiens.
XX
PN   WO9012803-A.
XX
PD   01-NOV-1990.
XX
PF   14-APR-1989;   89US-00338991.
XX
PR   14-APR-1989;   89US-00338991.
XX
PA   (BIOJ ) BIOGEN INC.
PA   (BIOJ ) BIOGEN INC.
XX
PI   Staples MA,  Pargellis CA;
XX
DR   WPI; 1990-348421/46.
XX
PT   Purifying protein having surface metal-binding amino acid residues -
PT   using an immobilised metal affinity chromatography resin.
XX
PS   Disclosure; Fig 2; 36pp; E.
XX
CC   The rs T4 can be purified from a crude sample (Updated on 25-MAR-2003 to
CC   correct PA field.)
XX
SQ   Sequence 375 AA;

  Query Match             81.9%;  Score 1904;  DB 2;  Length 375;
  Best Local Similarity   98.9%;  Pred. No. 6.7e-122;
  Matches  369;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         26 NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS 85
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS 60

Qy         86 LWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLES 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLES 120

Qy        146 PPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQ 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQ 180

Qy        206 KASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKR 265
              |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db        181 KASSIVYKKEGEQVEFSFPLAFTVQKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKR 240

Qy        266 VTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKN 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKN 300

Qy        326 LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIK 385
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        301 LTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIK 360

Qy        386 VLPTWSTPVHPRA 398
              ||||||||| | |
Db        361 VLPTWSTPVQPMA 373


RESULT 77
AAY54500
ID   AAY54500 standard; protein; 433 AA.
```

```
XX
AC   AAY54500;
XX
DT   25-APR-2000  (first entry)
XX
DE   Amino acid sequence of the human CD4 protein.
XX
KW   Human; CD4 protein; antigenic peptide; CDR2-like domain; apoptosis;
KW   syncytia formation; human immune deficiency virus; HIV binding;
KW   CD4-Class II interaction; immunisation; CD4 surface complex;
KW   immune response; transplant rejection; autoimmune disease;
KW   rheumatoid arthritis; systemic lupus erythematosus; psoriasis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          27. .66
FT                   /note= "CDR-2 like domain; specifically claimed in claim
FT                   1"
XX
PN   WO9967294-A1.
XX
PD   29-DEC-1999.
XX
PF   21-JUN-1999;   99WO-US014030.
XX
PR   20-JUN-1998;   98US-00100409.
XX
PA   (UNBI-) UNITED BIOMEDICAL INC.
XX
PI   Wang CY;
XX
DR   WPI; 2000-160579/14.
XX
PT   New antigenic peptide from the CDR2 domain of CD4, for immunization
PT   against e.g. human immune deficiency virus.
XX
PS   Claim 1; Page 70-71; 106pp; English.
XX
CC   The present sequence represents the human CD4 protein. The specification
CC   describes antigenic peptides derived from the CDR2-like domain of CD4
CC   (amino acids 27-66 of AAY54500). These antigenic peptides present
CC   neutralising receptor/co-receptor effector sites of the CDR2-like domain.
CC   The peptides evoke effective antibody responses by having optimised site-
CC   specificity. The induced antibodies block human immune deficiency virus
CC   (HIV) binding and syncytia formation. They may also block CD4-Class II
CC   interactions with other cells, deliver signals to T cells (inhibiting
CC   normal CD4+-mediated immunoregulatory functions) or induce apoptosis of
CC   CD4 cells by simultaneous engagement of T cell receptors. Conjugates and
CC   peptides containing the antigenic peptides are used for active
CC   immunisation to generate antibodies against CD4 surface complexes,
CC   especially to prevent binding of HIV to CD4 and thus HIV infection, but
CC   also to treat undesirable immune responses such as transplant rejection,
CC   or autoimmune diseases (rheumatoid arthritis, systemic lupus
CC   erythematosus or psoriasis). These conjugates produce high-titre
CC   antibodies which are broadly neutralising against primary isolates from
CC   all classes of HIV-1 and of HIV-2. The peptides may be cyclically
CC   constrained and may include a promiscuous T helper epitope that is active
CC   in genetically diverse subjects
XX
SQ   Sequence 433 AA;

  Query Match             81.9%;  Score 1904;  DB 3;  Length 433;
  Best Local Similarity   88.6%;  Pred. No. 8e-122;
  Matches  379;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIIFDLKNKEVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSSQVLLESNIKV 361

Qy        387 LPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGV- 445
              |||||||| |                                ||      |    || :|||:
Db        362 LPTWSTPVQP------------------------------MALIVLGGVAGLLLFIGLGIF 392

Qy        446 ACVLARTR 453
               ||  | |
Db        393 FCVRCRHR 400


RESULT 78
AAP96151
ID   AAP96151 standard; protein; 434 AA.
XX
AC   AAP96151;
XX
DT   25-MAR-2003  (revised)
DT   24-DEC-1990  (first entry)
XX
DE   Sequence of a fusion of the herpes virus and N-terminal 27 residues to
DE   the putative mature N-terminus of CD4T.
XX
KW   gp120; HIV; therapy; adheson; antiviral.
XX
OS   Synthetic.
XX
PN   EP314317-A.
XX
PD   03-MAY-1989.
XX
PF   03-OCT-1988;   88EP-00309194.
XX
PR   02-OCT-1987;   87US-00104329.
PR   28-SEP-1988;   88US-00250785.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Capon DJ,  Gregrory TJ;
XX
DR   WPI; 1989-131855/18.
DR   N-PSDB; AAN90778.
XX
PT   Compsns. contg. adheson variants - useful in therapy and diagnostics,
PT   e.g. CD4 variants which are therapeutically useful for treating human
PT   immuno-deficiency virus.
XX
PS   Example; Fig 2a-2c; 36pp; English.
XX
CC   A nucleic acid encoding an AA sequence variant of an adheson is claimed.
CC   Adheson variants may be a fusion of a CD4 polypeptide and a polypeptide
CC   different from CD4. Used therapeutically or as diagnostic reagents for
CC   the assay of adhesons or their ligands, or for purificn. of the adhesons.
CC   The CD4 adheson variants are used for antiviral or immunomodulatory
CC   therapy, in partic. treatment of HIV infection. (Updated on 25-MAR-2003
CC   to correct PI field.)
XX
SQ   Sequence 434 AA;
```

```
  Query Match           81.9%;  Score 1904;  DB 1;  Length 434;
  Best Local Similarity  99.7%;  Pred. No. 8e-122;
  Matches  369;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         56 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 115

Qy         84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 175

Qy        144 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 235

Qy        204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 295

Qy        264 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 355

Qy        324 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESN 383
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        356 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESN 415

Qy        384 IKVLPTWSTP 393
              ||||||||||
Db        416 IKVLPTWSTP 425


RESULT 79
AAP93557
ID   AAP93557 standard; protein; 434 AA.
XX
AC   AAP93557;
XX
DT   25-MAR-2003  (revised)
DT   03-OCT-2002  (revised)
DT   06-JUN-1990  (first entry)
XX
DE   Fusion of the herpes gD leader and N-terminal 27 residues to mature N-
DE   terminus of CD4T.
XX
KW   Adheson variant; CD4T; soluble CD4; truncated CD4; antiviral; HIV;
KW   herpes gD; immunomodulatory; diagnostic.
XX
OS   Homo sapiens.
XX
PN   WO8902922-A.
XX
PD   06-APR-1989.
XX
PF   03-OCT-1988;   88WO-US003414.
XX
PR   02-OCT-1987;   87US-00104329.
PR   28-SEP-1988;   88US-00250785.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Capon DJ,  Gregory TJ;
XX
DR   WPI; 1989-114397/15.
DR   N-PSDB; AAN90735.
XX
PT   New nucleic acid sequences encoding adheson, esp. CD 4, variants -
PT   partic. with trans-membrane domain inactivated or fused to other peptide,
PT   useful esp. for treating HIV infections.
XX
PS   Fig 2A-2C; pp. 5/13-8/13; 78pp; English.
XX
CC   CD4T is a truncated or soluble variant of CD4. CD4T fusion proteins can
CC   have antiviral and immunomodulatory activity and are esp. useful for
CC   treating HIV infections, regardless of genetic variation within the
CC   virus. CD4T fusion proteins, and antibodies raised against them, can also
CC   be used diagnostically for assaying adhesons and their ligands. (Updated
CC   on 03-OCT-2002 to add missing OS field.) (Updated on 25-MAR-2003 to
CC   correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 434 AA;

  Query Match           81.9%;  Score 1904;  DB 1;  Length 434;
  Best Local Similarity  99.7%;  Pred. No. 8e-122;
  Matches  369;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         56 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 115

Qy         84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 175

Qy        144 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 235

Qy        204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 295

Qy        264 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 355

Qy        324 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESN 383
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        356 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESN 415

Qy        384 IKVLPTWSTP 393
              ||||||||||
Db        416 IKVLPTWSTP 425


RESULT 80
AAR74222
ID   AAR74222 standard; protein; 432 AA.
XX
AC   AAR74222;
XX
DT   25-MAR-2003  (revised)
DT   26-NOV-1995  (first entry)
XX
DE   Epitope on the primary CD4 sequence.
XX
KW   Chimaeric; mutant; mapping; immunodiagnostics.
XX
OS   Synthetic.
XX
PN   US5411861-A.
XX
PD   02-MAY-1995.
XX
PF   27-FEB-1992;   92US-00842465.
XX
PR   15-APR-1988;   88US-00181826.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Peterson A;
XX
DR   WPI; 1995-178122/23.
XX
```

```
PT   Mutational analysis method for protein epitope(s) - by expressing mutant
PT   cDNA and using negative and positive selection to identify binding loss
PT   mutants.
XX
PS   Disclosure; Fig 2; 28pp; English.
XX
CC   The sequence is that of an epitope on the primary CD4 sequence. The
CC   epitope sequence can be identified by vector piH3M. The See also
CC   AAR74221. (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 432 AA;

  Query Match             81.8%;  Score 1901;  DB 2;  Length 432;
  Best Local Similarity   88.6%;  Pred. No. 1.3e-121;
  Matches  379;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              |||||||||||||||||||||||||||| ||||||||||||||| |||||||||||||
Db          1 KVVLGKKGDTVELTCTASQKKSIQFHWKNWNQIKILGNQGSFLTKQPSKLNDRADSRRSL 60

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 120

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 180

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 240

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 300

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 360

Qy        387 LPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGV- 445
              |||||||| |                                ||      |   || :|||:
Db        361 LPTWSTPVQP----------------------------MALIVLGGVAGLLLFIGLGIF 391

Qy        446 ACVLARTR 453
              ||  | |
Db        392 FCVRCRHR 399


RESULT 81
AAY30514
ID   AAY30514 standard; protein; 432 AA.
XX
AC   AAY30514;
XX
DT   15-NOV-1999  (first entry)
XX
DE   Predicted sequence of the CD4 protein.
XX
KW   CD4 protein; rapid mutational analysis method; protein epitope mapping;
KW   binding domain mapping; binding capacity; anti-CD2 antibody;
KW   anti-CD4 antibody; ligand binding site study.
XX
OS   Homo sapiens.
XX
PN   US5955264-A.
XX
PD   21-SEP-1999.
XX
PF   11-OCT-1994;   94US-00320663.
XX
PR   15-APR-1988;   88US-00181826.
PR   27-FEB-1992;   92US-00842465.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Peterson A;
XX
DR   WPI; 1999-550602/46.
XX
PT   Rapid mutational analysis method for mapping protein epitopes.
XX
PS   Example 4; Fig 4; 27pp; English.
XX
CC   The present sequence represents the predicted sequence of the CD4
CC   protein. The protein is used to demonstrate the method of the invention.
CC   The specification describes a rapid mutational analysis method for
CC   mapping protein epitopes and binding domains, by identifying substitution
CC   mutations that result in the loss of binding capacity. The method may be
CC   used for mapping protein epitopes, antigenic domains and binding sites.
CC   It has been used for mapping binding sites for sixteen anti-CD2 and anti-
CC   CD4 monoclonal antibodies. The method is especially useful for ligand
CC   binding site studies for the design of new ligands and drugs
XX
SQ   Sequence 432 AA;

  Query Match             81.8%;  Score 1901;  DB 2;  Length 432;
  Best Local Similarity   88.6%;  Pred. No. 1.3e-121;
  Matches  379;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              |||||||||||||||||||||||||||| ||||||||||||||| |||||||||||||
Db          1 KVVLGKKGDTVELTCTASQKKSIQFHWKNWNQIKILGNQGSFLTKQPSKLNDRADSRRSL 60

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 120

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 180

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 240

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 300

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 360

Qy        387 LPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGV- 445
              |||||||| |                                ||      |   || :|||:
Db        361 LPTWSTPVQP----------------------------MALIVLGGVAGLLLFIGLGIF 391

Qy        446 ACVLARTR 453
              ||  | |
Db        392 FCVRCRHR 399


RESULT 82
ADA25188
ID   ADA25188 standard; peptide; 432 AA.
XX
AC   ADA25188;
XX
DT   20-NOV-2003  (first entry)
XX
DE   CD4 epitope region 1.
XX
KW   ligand binding site study; ligand design; drug design; CD4; epitope;
KW   immunogenic.
XX
```

```
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   Region          34. .52
FT                   /note= "Leu3a epitope"
FT   Region          364. .393
FT                   /note= "Transmembrane region"
XX
PN   US6579676-B1.
XX
PD   17-JUN-2003.
XX
PF   21-SEP-1999;   99US-00400207.
XX
PR   15-APR-1988;   88US-00181826.
PR   27-FEB-1992;   92US-00842465.
PR   11-OCT-1994;   94US-00320663.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Seed B,  Peterson A;
XX
DR   WPI; 2003-644358/61.
XX
PT   New population of recombinant cells collectively comprising mutant cDNAs,
PT   useful in ligand binding site studies for the design of new ligands and
PT   drugs.
XX
PS   Disclosure; Fig 4; 27pp; English.
XX
CC   The invention relates to a population of recombinant cells collectively
CC   comprising mutant cDNAs comprising substitution mutants, where each
CC   mutant cDNA encodes a particular protein of interest, where the
CC   nucleotide sequence encoding the protein is known and the protein of
CC   interest comprises at least two known binding domains comprising
CC   different ligands. The population of recombinant cells collectively
CC   comprising mutant cDNAs are useful in ligand binding site studies for the
CC   design of new ligands and drugs. The present sequence represents the
CC   amino acid sequence of the CD4 epitope region 1.
XX
SQ   Sequence 432 AA;

  Query Match             81.8%;  Score 1901;  DB 6;  Length 432;
  Best Local Similarity   88.6%;  Pred. No. 1.3e-121;
  Matches  379;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              |||||||||||||||||||||||||| ||||||||||||| ||||||||||||||
Db          1 KVVLGKKGDTVELTCTASQKKSIQFHWKNWNQIKILGNQGSFLTKQPSKLNDRADSRRSL 60

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 120

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 180

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 240

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 300

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 360

Qy        387 LPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGV- 445
              |||||||| |                                ||     |   || :|||:
Db        361 LPTWSTPVQP----------------------------MALIVLGGVAGLLLFIGLGIF 391

Qy        446 ACVLARTR 453
                ||  | |
Db        392 FCVRCRHR 399


RESULT 83
AAP93528
ID   AAP93528 standard; protein; 370 AA.
XX
AC   AAP93528;
XX
DT   25-MAR-2003  (revised)
DT   04-JUN-1990  (first entry)
XX
DE   Human soluble CD4 protein (T4ex1) encoded by T4 SEC1 cDNA.
XX
KW   Human soluble CD4 protein; T4 SEC1 cDNA; T4ex1;
KW   HIV gp120 envelope protein; T-lymphocyte.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 129
FT   Misc-difference 163
FT   Misc-difference 183
FT   Misc-difference 243
FT   Misc-difference 370
XX
PN   WO8903222-A.
XX
PD   20-APR-1989.
XX
PF   05-OCT-1988;   88WO-US003454.
XX
PR   08-OCT-1987;   87US-00106185.
PR   14-JAN-1988;   88US-00144313.
PR   14-JUN-1988;   88US-00206585.
PR   11-JUL-1988;   88US-00217475.
XX
PA   (DANA-) DANA-FABER CANCER.
XX
PI   Reinherz E,  Hussey R,  Droski J,  Richardson N;
XX
DR   WPI; 1989-129898/17.
DR   N-PSDB; AAN90764.
XX
PT   Soluble human CD4 fragments - capable of binding HIV gp 120 envelope
PT   protein, used for diagnosis, treatment or prophylaxis of HIV infection.
XX
PS   Fig 1; pages 1/11-3/11; 106pp; English.
XX
CC   Misc-difference in the features table represent where a modification in
CC   cDNA and in the encoded CD4 protein has been made. A soluble human CD4
CC   fragment capable of binding to the gp. 120 envelope protein of HIV is
CC   claimed. Such a fragment should include none of the hydrophobic trans-
CC   membrane region of naturally-occurring CD4, or contain a portion of the
CC   hydrophobic region which is sufficiently short it does not prevent
CC   solubilisation. Also claimed are modified soluble CD4 fragments which
CC   differ by substn. or addn. to this sequence (AAP93528) by at least one
CC   AA. Such fragments pref. consist of: the first 182, 369, 549, 729 or 1107
CC   AAs, domain 1,2, and partial domain 3-pref. where the protein is
CC   truncated at AA posn. 243; Asn-271 is Asp and/or Asn-300 is Asp; or the
CC   AA sequence without the N-linked glycosylation sites, of CD4 protein.
CC   Such fragments can also contain the corresp. AA of murine CD4 substd. at
CC   more than one triplet site in the DNA, selected from a triplet encoding
CC   the AA at posn. 48; 50; 51; 121; 122; 123; 155; 156; or 158 of human CD4.
CC   CD4 fragments have the capacity to prevent infection of human T-
CC   lymphocyte infection by HIV and to prevent the formation of human T-
CC   lymphocyte syncytia. They can be used in the diagnosis, treatment and
CC   prevention of HIV infection. (Updated on 25-MAR-2003 to correct PR
CC   field.) (Updated on 25-MAR-2003 to correct PI field.)
XX
```

```
SQ   Sequence 370 AA;

  Query Match             81.7%;  Score 1899;  DB 1;  Length 370;
  Best Local Similarity   99.7%;  Pred. No. 1.4e-121;
  Matches  368;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 84
AAB83356
ID   AAB83356 standard; protein; 370 AA.
XX
AC   AAB83356;
XX
DT   09-OCT-2001  (first entry)
XX
DE   Human CD4 protein sequence.
XX
KW   Chemotactic chemokine receptor 5; gp120; CD4; therapy; HIV; CCR5;
KW   human immunodeficiency virus; anti-inflammatory disease; human.
XX
OS   Homo sapiens.
XX
PN   EP1118858-A2.
XX
PD   25-JUL-2001.
XX
PF   03-JAN-2001; 2001EP-00300020.
XX
PR   12-JAN-2000; 2000GB-00000659.
PR   12-JAN-2000; 2000GB-00000661.
PR   12-JAN-2000; 2000GB-00000663.
XX
PA   (PFIZ ) PFIZER LTD.
PA   (PFIZ ) PFIZER INC.
XX
PI   Dobbs S,  Perros M,  Rickett GA;
XX
DR   WPI; 2001-477088/52.
DR   N-PSDB; AAF87101.
XX
PT   Determining if an agent can modulate CCR5-gp120 interaction, comprises
PT   incubating the agent with CCR5 and gp120 and determining if the agent
PT   modulates the interaction.
XX
PS   Claim 1; Page 112; 113pp; English.
XX
CC   This sequence represents the human CD4 protein sequence. The invention
CC   relates to a method for determining whether an agent is capable of
CC   modulating the interaction of chemotactic chemokine receptor 5 (CCR5)
CC   with gp120, comprising incubating the agent with CCR5 and gp120 and
CC   determining whether the agent modulates the interaction, where gp120 is
CC   associated with CD4, and where the interaction is a low affinity binding.
CC   The method is used to identify an agent capable of modulating the
CC   interaction of CCR5 with gp120. An agent identified by the method is used
CC   to prepare a pharmaceutical composition for the treatment of a disease or
CC   condition associated with CCR5 and gp120 interaction, to treat a subject
CC   with a disease or condition associated with CCR5 and gp120 interaction,
CC   and for preparing a pharmaceutical for treating human immunodeficiency
CC   virus (HIV). It can also be treat anti-inflammatory diseases. The method
CC   is commercially useful, amenable to high throughput screening, and
CC   detects interaction of gp120 with cells expressing only CCR5
XX
SQ   Sequence 370 AA;

  Query Match             81.7%;  Score 1899;  DB 4;  Length 370;
  Best Local Similarity   99.7%;  Pred. No. 1.4e-121;
  Matches  368;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 85
AAY39824
ID   AAY39824 standard; protein; 369 AA.
XX
AC   AAY39824;
XX
DT   03-DEC-1999  (first entry)
XX
DE   Soluble human T4 protein.
XX
KW   Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;
KW   vaccine; immunisation; therapy.
XX
OS   Homo sapiens.
XX
PN   US5958678-A.
XX
PD   28-SEP-1999.
XX
PF   12-DEC-1994;   94US-00354452.
XX
```

```
PR   21-AUG-1986;   86US-00898587.
PR   11-JUN-1991;   91US-00713564.
PR   06-JUL-1992;   92US-00909021.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Mcdougal JS,  Weiss R,  Axel R,  Littman DR,  Maddon PJ,  Chess L;
XX
DR   WPI; 1999-561025/47.
XX
PT   Human T4 protein inhibits HIV binding to T4 cells, useful for treating
PT   AIDS.
XX
PS   Claim 1; Col 51-53; 58pp; English.
XX
CC   This sequence represents the soluble human T4 protein of the invention.
CC   The soluble human T4 protein blocks the binding of HIV to T4+ cells and
CC   is therefore useful for the treatment of AIDS. Monoclonal antibodies
CC   against the T4 protein may be used as vaccines for immunising subjects
CC   against AIDS
XX
SQ   Sequence 369 AA;

  Query Match             81.3%;  Score 1891;  DB 2;  Length 369;
  Best Local Similarity   99.7%;  Pred. No. 5.1e-121;
  Matches  367;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPV 394
              ||||||||
Db        362 LPTWSTPV 369


RESULT 86
AAY88327
ID   AAY88327 standard; protein; 369 AA.
XX
AC   AAY88327;
XX
DT   14-JUL-2000  (first entry)
XX
DE   T4 glycoprotein amino acid sequence.
XX
KW   sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;
KW   AIDS; treatment; inhibit; cell to cell spread; infection; fusion.
XX
OS   Mammalia.
XX
PN   US5126433-A.
XX
PD   30-JUN-1992.
XX
PF   23-OCT-1987;   87US-00114244.
XX
PR   21-AUG-1986;   86US-00898587.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Maddon PJ,  Chess L,  Axel R,  Weiss R,  Littman DR,  Mcdougal JS;
XX
DR   WPI; 2000-348913/30.
XX
PT   Soluble T4 glycoprotein useful for prevention and treatment of acquired
PT   immunodeficiency syndrome and for screening inhibitors of human
PT   immunodeficiency viral binding.
XX
PS   Claim 1; Col 54; 64pp; English.
XX
CC   This sequence represents the amino acid sequence of glycosylated sT4
CC   glycoprotein. Human immunodeficiency virus (HIV) uses sT4 as a target
CC   receptor on T cells. The invention relates to glycosylated sT4 which
CC   functions by blocking the binding of HIV to T4 target cells, and can be
CC   used for the prophylaxis and treatment of AIDS patients. Administration
CC   of sT4 effectively inhibits the cell to cell spreading of HIV infection
CC   and also the fusion of HIV-infected T4 cells and non-infected T4 cells.
CC   The administration of T4 alleviates several symptoms associated with
CC   AIDS, and prevents the occurrence of new pathological changes. The sT4
CC   glycoprotein is useful for the prophylaxis and treatment of patients with
CC   AIDS. It is also useful as a reagent to identify natural, synthetic or
CC   recombinant molecules which act as therapeutic agents or inhibitors of
CC   T4+ cell interactions and in diagnostic assays for detection T4 proteins
CC   or molecules
XX
SQ   Sequence 369 AA;

  Query Match             81.3%;  Score 1891;  DB 3;  Length 369;
  Best Local Similarity   99.7%;  Pred. No. 5.1e-121;
  Matches  367;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPV 394
              ||||||||
Db        362 LPTWSTPV 369


RESULT 87
AAR12956
ID   AAR12956 standard; protein; 370 AA.
XX
AC   AAR12956;
```

```
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Lys(64) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Disclosure; Table 1; 41pp; English.
XX
CC   Mutant M4 is one of 17 modified CD4 proteins analysed for class II MHC
CC   binding. It corresponds to domains I and II of the human CD4 protein in
CC   which Gln at position 64 is replaced with Lys. Mutant M4 is not one of
CC   the preferred mutants of the invention as it retains its Class II MHC
CC   binding affinity. See AAR12951-R12967. (Updated on 09-JAN-2003 to add
CC   missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 370 AA;

  Query Match             81.2%;  Score 1887;  DB 2;  Length 370;
  Best Local Similarity   99.2%;  Pred. No. 9.6e-121;
  Matches  366;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDKGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 88
AAR12954
ID   AAR12954 standard; protein; 370 AA.
XX
AC   AAR12954;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   His(40) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Disclosure; Table 1; 41pp; English.
XX
CC   Mutant M2 is one of 17 modified CD4 proteins analysed for class II MHC
CC   binding. It corresponds to domains I and II of the human CD4 protein in
CC   which Gln at position 40 is replaced with His. Mutant M2 is not one of
CC   the preferred mutants of the invention as it retains its Class II MHC
CC   binding affinity. See AAR12951-R12967. (Updated on 09-JAN-2003 to add
CC   missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 370 AA;

  Query Match             81.1%;  Score 1886;  DB 2;  Length 370;
  Best Local Similarity   99.2%;  Pred. No. 1.1e-120;
  Matches  366;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNHGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 89
AAR12964
ID   AAR12964 standard; protein; 370 AA.
XX
AC   AAR12964;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Val(143) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Disclosure; Table 1; 41pp; English.
XX
CC   Mutant M12 is one of 17 modified CD4 proteins analysed for class II MHC
CC   binding. It corresponds to domains I and II of the human CD4 protein in
CC   which Thr at position 143 is replaced with Val. Mutant M12 is not one of
CC   the preferred mutants of the invention as it retains its Class II MHC
CC   binding affinity. See AAR12951-R12967. (Updated on 09-JAN-2003 to add
CC   missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 370 AA;

  Query Match             81.1%;  Score 1886;  DB 2;  Length 370;
  Best Local Similarity   99.2%;  Pred. No. 1.1e-120;
  Matches  366;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKVLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 90
AAR12965
ID   AAR12965 standard; protein; 370 AA.
XX
AC   AAR12965;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Arg(150) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Disclosure; Table 1; 41pp; English.
XX
CC   Mutant M13 is one of 17 modified CD4 proteins analysed for class II MHC
CC   binding. It corresponds to domains I and II of the human CD4 protein in
CC   which Glu at position 150 is replaced with Arg. Mutant M13 is not one of
CC   the preferred mutants of the invention as it retains its Class II MHC
CC   binding affinity. See AAR12951-R12967. (Updated on 09-JAN-2003 to add
CC   missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 370 AA;

  Query Match             81.1%;  Score 1886;  DB 2;  Length 370;
  Best Local Similarity   99.2%;  Pred. No. 1.1e-120;
  Matches  366;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLRLQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 91
AAR12958
ID   AAR12958 standard; protein; 370 AA.
XX
AC   AAR12958;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Gln(80) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Disclosure; Table 1; 41pp; English.
XX
CC   Mutant M6 is one of 17 modified CD4 proteins analysed for class II MHC
CC   binding. It corresponds to domains I and II of the human CD4 protein in
CC   which Asp at position 80 is replaced with Gln. Mutant M6 is not one of
CC   the preferred mutants of the invention as it could not be evaluated for
CC   Class II MHC binding affinity. The substitutions grossly affect the
CC   structure of the external CD4 domains so that reactivity with all anti-
CC   CD4 monclonal antibodies tested was reduced. See AAR12951-R12967.
CC   (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-MAR-2003
CC   to correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 370 AA;

  Query Match             81.1%;  Score 1885;  DB 2;  Length 370;
  Best Local Similarity   99.2%;  Pred. No. 1.3e-120;
  Matches  366;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSQTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 92
AAR12951
ID   AAR12951 standard; protein; 370 AA.
XX
AC   AAR12951;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Glu(17), Ser(18) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
```

```
PS   Claim 3; Table 1; 41pp; English.
XX
CC   Mutant M1.1 is one example of a novel CD4 protein of the invention. It
CC   corresponds to domains I and II of the human CD4 protein in which Thr at
CC   position 17 and Ala at position 18 are replaced with Glu and Ser,
CC   respectively. The modified CD4 can inhibit HIV binding to the CD4
CC   receptors of T lymphocytes and deplete circulating levels of HIV gp120;
CC   it does not bind class II antigen bearing B lymphocytes. See also
CC   AAR12952-R12967. (Updated on 09-JAN-2003 to add missing OS field.)
CC   (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
CC   correct PA field.)
XX
SQ   Sequence 370 AA;

  Query Match             80.9%;  Score 1882;  DB 2;  Length 370;
  Best Local Similarity   98.9%;  Pred. No. 2.1e-120;
  Matches  365;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||| :|||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCESSQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 93
AAR12957
ID   AAR12957 standard; protein; 370 AA.
XX
AC   AAR12957;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Asn(72), Lys(73) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Claim 3; Table 1; 41pp; English.
XX
CC   Mutant M5 is one example of a novel CD4 protein of the invention. It
CC   corresponds to domains I and II of the human CD4 protein in which Lys at
CC   position 72 and Asn at position 73 are replaced with Asn and Lys,
CC   respectively. The modified CD4 can inhibit HIV binding to the CD4
CC   receptors of T lymphocytes and deplete circulating levels of HIV gp120;
CC   it does not bind class II antigen bearing B lymphocytes. See AAR12951-
CC   R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-
CC   MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA
CC   field.)
XX
SQ   Sequence 370 AA;

  Query Match             80.9%;  Score 1880;  DB 2;  Length 370;
  Best Local Similarity   98.9%;  Pred. No. 2.9e-120;
  Matches  365;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIINKLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 94
AAR12959
ID   AAR12959 standard; protein; 370 AA.
XX
AC   AAR12959;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Asn(88), Arg(89), Glu(94) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
```

```
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Disclosure; Table 1; 41pp; English.
XX
CC   Mutant M7 is one of 17 modified CD4 proteins analysed for class II MHC
CC   binding. It corresponds to domains I and II of the human CD4 protein in
CC   which Asp(88), Gln(89) and Gln(94) are replaced with Asn, Arg and Glu,
CC   respectively. Mutant M7 is not one of the preferred mutants of the
CC   invention as it retains its Class II MHC binding affinity. See AAR12951-
CC   R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-
CC   MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA
CC   field.)
XX
SQ   Sequence 370 AA;

  Query Match             80.8%;  Score 1879;  DB 2;  Length 370;
  Best Local Similarity   98.6%;  Pred. No. 3.4e-120;
  Matches  364;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||::|||||:|||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVENRKEEVELLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 95
AAR12962
ID   AAR12962 standard; protein; 370 AA.
XX
AC   AAR12962;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Leu(127), Thr(128), Glu(129) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Disclosure; Table 1; 41pp; English.
XX
CC   Mutant M10 is one of 17 modified CD4 proteins analysed for class II MHC
CC   binding. It corresponds to domains I and II of the human CD4 protein in
CC   which Ser(127), Val(128) and Gln(129) are replaced with Leu, Thr and Glu,
CC   respectively. Mutant M10 is not one of the preferred mutants of the
CC   invention as it retains its Class II MHC binding affinity. See AAR12951-
CC   R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-
CC   MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA
CC   field.)
XX
SQ   Sequence 370 AA;

  Query Match             80.8%;  Score 1879;  DB 2;  Length 370;
  Best Local Similarity   98.6%;  Pred. No. 3.4e-120;
  Matches  364;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              |||||  :|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPHTECRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370
```

```
RESULT 96
AAP90992
ID   AAP90992 standard; peptide; 435 AA.
XX
AC   AAP90992;
XX
DT   25-MAR-2003  (revised)
DT   04-JUN-1990  (first entry)
XX
DE   Human CD4 antigen.
XX
KW   Oligopeptide; human CD4 antigen; retrovirus; HIV; therapy; HIV-1; HIV-2;
KW   proplylactic; vaccine; diagnostic agent.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          83. .89
FT                   /note= "Core sequence"
XX
PN   WO8903420-A.
XX
PD   20-APR-1989.
XX
PF   04-AUG-1988;   88WO-US002638.
XX
PR   13-OCT-1987;   87US-00108160.
PR   01-JUN-1988;   88US-00203285.
PR   13-OCT-1988;   88WO-US003592.
XX
PA   (GENE-) GENELABS INC.
XX
PI   Lifson JD,  Hwang K,  Eiden LE,  Nara PL,  Fraser B;
XX
DR   WPI; 1989-130038/17.
XX
PT   Oligopeptide cpds. having partial sequence of human cd4 antigen -
PT   inhibiting interaction between CD4 positive cells and retrovirus, esp.
PT   HIV.
XX
PS   Disclosure; Page 6; 62pp; English.
XX
CC   The patent concerns oligopeptides derived from the sequence as described
CC   in the patent title. Compositions of interest will include a core
CC   sequence (see Features Table), which may be alone, or more usually will
CC   comprise at least one additional amino acid on at least one end of the
CC   core sequence. They will contain more than one heteroatom of a bond (not
CC   a peptide bond) of more than one amino acid joined with a derivatising
CC   gp. In the preferred sequence, the heteroatom is S and 86-Cys is
CC   derivatised, and/or the heteroatom is O and 87-Glu is derivatised. The
CC   derivatising gp. is 7-12C aralkyl opt. contg. 1-2 heteroatoms, esp.
CC   benzyl or chlorobenzyl. The oligopeptides can be used therapeutically,
CC   prophylactically or in vitro to modulate response to retroviruses, esp.
CC   HIV-1 or HIV-2, esp. the fusion of CD4-bearing cells or virion
CC   infectivity. They are also potentially useful in vaccine development and
CC   to produce antibodies which are useful as antiviral agents or diagnostic
CC   reagents. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-
CC   MAR-2003 to correct PI field.)
XX
SQ   Sequence 435 AA;

  Query Match             80.8%;  Score 1878;  DB 1;  Length 435;
  Best Local Similarity   87.9%;  Pred. No. 4.8e-120;
  Matches  379;  Conservative    2;  Mismatches   20;  Indels   30;  Gaps    2;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 60

Qy         84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 143
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RSLTDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 120

Qy        144 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 180

Qy        204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 240

Qy        264 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 323
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        241 KRVTQDPKLQMGKKLPLHLTLPQALPQYAQSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 300

Qy        324 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESN 383
              ||||||| |||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db        301 KNLTCEVTGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMTQCLLSDSGQVLLESN 360

Qy        384 IKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGL 443
              |||||||||| |                                    ||     |   || :||
Db        361 IKVLPTWSTPVQP----------------------------MALIVLGGVAGLLLFIGL 391

Qy        444 GV-ACVLARTR 453
              |: || | |
Db        392 GIFFCVRCRHR 402


RESULT 97
AAR12966
ID   AAR12966 standard; protein; 370 AA.
XX
AC   AAR12966;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Thr(162), Leu(163), Asp(164) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Disclosure; Table 1; 41pp; English.
XX
CC   Mutant M15 is one of 17 modified CD4 proteins analysed for class II MHC
CC   binding. It corresponds to domains I and II of the human CD4 protein in
CC   which Leu(162), Gln(163) and Asn(164) are replaced with Thr, Leu and Asp,
CC   respectively. Mutant M15 is not one of the preferred mutants of the
CC   invention as it retains its Class II MHC binding affinity. See AAR12951-
CC   R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-
CC   MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA
CC   field.)
```

```
XX
SQ   Sequence 370 AA;

  Query Match             80.6%;  Score 1874;  DB 2;  Length 370;
  Best Local Similarity   98.6%;  Pred. No. 7.4e-120;
  Matches  364;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||  :|||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVTLDQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 98
AAR12952
ID   AAR12952 standard; protein; 370 AA.
XX
AC   AAR12952;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Ile(23), Thr(24), Val(25) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Claim 3; Table 1; 41pp; English.
XX
CC   Mutant M1.2 is one example of a novel CD4 protein of the invention. It
CC   corresponds to domains I and II of the human CD4 protein in which
CC   Ser(23), Ile(24) and Gln(25) are replaced with Ile, Thr and Val,
CC   respectively. The modified CD4 can inhibit HIV binding to the CD4
CC   receptors of T lymphocytes and deplete circulating levels of HIV gp120;
CC   it does not bind class II antigen bearing B lymphocytes. See AAR12951-
CC   R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-
CC   MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA
CC   field.)
XX
SQ   Sequence 370 AA;

  Query Match             80.6%;  Score 1873;  DB 2;  Length 370;
  Best Local Similarity   98.6%;  Pred. No. 8.7e-120;
  Matches  364;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKITVFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 99
AAR12967
ID   AAR12967 standard; protein; 370 AA.
XX
AC   AAR12967;
XX
DT   25-MAR-2003  (revised)
DT   09-JAN-2003  (revised)
DT   18-SEP-1991  (first entry)
XX
DE   Asp(155), Phe(156), Asn(158) CD4 mutant.
XX
KW   CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW   human immunodeficiency virus.
XX
OS   Homo sapiens.
XX
PN   WO9109123-A.
XX
PD   27-JUN-1991.
XX
PF   15-DEC-1989;   89WO-US005625.
XX
PR   15-DEC-1989;   89WO-US005625.
XX
PA   (DAND ) DANA FARBER CANCER INST INC.
PA   (REIN/) REINHERZ E L.
```

```
XX
PI   Reinherz E,  Clayton LK;
XX
DR   WPI; 1991-208149/28.
XX
PT   New soluble human CD4 glyco:protein - has decreased affinity for MHC
PT   class II antigens and is used to treat, prevent and diagnose HIV
PT   infections.
XX
PS   Disclosure; Table 1; 41pp; English.
XX
CC   Mutant M14 is one of 17 modified CD4 proteins analysed for class II MHC
CC   binding. It corresponds to domains I and II of the human CD4 protein in
CC   which Gly(155), Thr(156) and Thr(158) are replaced with Asp, Phe and Asn,
CC   respectively. Mutant M14 is not one of the preferred mutants of the
CC   invention as it could not be evaluated for Class II MHC binding affinity.
CC   The substitutions grossly affect the structure of the external CD4
CC   domains so that reactivity with all anti-CD4 monclonal antibodies tested
CC   was eliminated. See AAR12951-R12966. (Updated on 09-JAN-2003 to add
CC   missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC   on 25-MAR-2003 to correct PA field.)
XX
SQ   Sequence 370 AA;

  Query Match             80.5%;  Score 1872;  DB 2;  Length 370;
  Best Local Similarity   98.6%;  Pred. No. 1e-119;
  Matches  364;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              |||||||||||||||||||||||||||||||  | ||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSDFWNCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKWVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 100
AAP91289
ID   AAP91289 standard; protein; 435 AA.
XX
AC   AAP91289;
XX
DT   25-MAR-2003  (revised)
DT   21-JAN-1991  (first entry)
XX
DE   Sequence of a segment of the human CD4 molecule.
XX
KW   HIV-1; HIV-2; vaccine; gp 120; diagnosis; retrovirus.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          83. .89
FT                   /note= "core peptide"
XX
PN   WO8903813-A.
XX
PD   05-MAY-1989.
XX
PF   13-OCT-1988;   88WO-US003592.
XX
PR   13-OCT-1987;   87US-00108160.
PR   01-JUN-1988;   88US-00203285.
XX
PA   (GENE-) GENELABS INC.
XX
PI   Lifson J,  Hwang K,  Eiden LE,  Nara PL,  Fracer B;
XX
DR   WPI; 1989-150731/20.
XX
PT   Anti-retro-viral polypeptide(s) - comprising a sequence of human CD4
PT   having amino acid joined to deriving gp.
XX
PS   Disclosure; Table 1, P 7; 67pp; English.
XX
CC   Polypeptide is claimed, comprising a sequence of 7 or more consecutive aa
CC   of human CD4. Also claimed is a cpd. comprising a sequence of 10 or more
CC   consecutive aa of human CD4 which includes the core peptide (see FT).
CC   (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
CC   correct PI field.)
XX
SQ   Sequence 435 AA;

  Query Match             80.5%;  Score 1871;  DB 1;  Length 435;
  Best Local Similarity   87.7%;  Pred. No. 1.4e-119;
  Matches  378;  Conservative    2;  Mismatches   21;  Indels   30;  Gaps    2;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 60

Qy         84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 143
              ||| |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db         61 RSLTDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLYANSDTHLLQGQSLTLTL 120

Qy        144 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 180

Qy        204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 240

Qy        264 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 323
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        241 KRVTQDPKLQMGKKLPLHLTLPQALPQYAQSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 300

Qy        324 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESN 383
              ||||||| |||||||||||||||||||||| |||||||||| ||||||||||||||||||
Db        301 KNLTCEVTGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMTQCLLSDSGQVLLESN 360

Qy        384 IKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGL 443
              |||||||||| |                                ||      |   || :||
Db        361 IKVLPTWSTPVQP----------------------------MALIVLGGVAGLLLFIGL 391

Qy        444 GV-ACVLARTR 453
              |: || | |
Db        392 GIFFCVRCRHR 402


Search completed: August  3, 2004, 13:08:17
Job time : 52.0872 secs
```

OM protein - protein search, using sw model

Run on: August 3, 2004, 13:14:00 ; Search time 38.3425 Seconds
(without alignments)
3706.029 Million cell updates/sec

Title: SEQ5
Perfect score: 2325
Sequence: 1 MNRGVPFRHLLLVLQLALLP.....VISFLLGLGLGVACVLARTR 453

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2040	87.7	462	10	US-09-939-537-5	Sequence 5, Appli
2	2040	87.7	462	11	US-09-243-008-5	Sequence 5, Appli
3	2038.5	87.7	575	10	US-09-939-537-4	Sequence 4, Appli
4	2038.5	87.7	575	11	US-09-243-008-4	Sequence 4, Appli
5	2038	87.7	457	11	US-09-891-119A-9	Sequence 9, Appli
6	2037	87.6	398	10	US-09-939-537-29	Sequence 29, Appl
7	2037	87.6	532	10	US-09-939-537-6	Sequence 6, Appli
8	2037	87.6	532	11	US-09-243-008-6	Sequence 6, Appli
9	2032	87.4	458	14	US-10-103-597A-39	Sequence 39, Appl
10	2032	87.4	458	14	US-10-188-444-39	Sequence 39, Appl
11	2032	87.4	458	14	US-10-207-655-170	Sequence 170, App
12	2024	87.1	458	12	US-10-151-274-3	Sequence 3, Appli
13	2017	86.8	402	14	US-10-097-044A-1	Sequence 1, Appli
14	2014	86.6	458	8	US-08-681-219-27	Sequence 27, Appl
15	2014	86.6	458	11	US-09-230-111C-25	Sequence 25, Appl
16	2014	86.6	458	14	US-10-092-138-25	Sequence 25, Appl
17	2001	86.1	397	11	US-09-891-119A-2	Sequence 2, Appli
18	1904	81.9	434	14	US-10-097-044A-4	Sequence 4, Appli
19	1899	81.7	370	9	US-09-759-841-6	Sequence 6, Appli
20	1894.5	81.5	448	14	US-10-024-329-32	Sequence 32, Appl
21	1050.5	45.2	310	8	US-08-485-163-7	Sequence 7, Appli
22	1050.5	45.2	310	9	US-09-766-995-6	Sequence 6, Appli
23	1046	45.0	530	8	US-08-485-163-5	Sequence 5, Appli
24	1046	45.0	530	9	US-09-766-995-4	Sequence 4, Appli
25	1045	44.9	432	8	US-08-485-163-3	Sequence 3, Appli
26	1045	44.9	432	9	US-09-766-995-2	Sequence 2, Appli
27	1041	44.8	203	10	US-09-939-537-31	Sequence 31, Appl
28	993	42.7	612	14	US-10-125-692-10	Sequence 10, Appl
29	921	39.6	788	14	US-10-073-118-26	Sequence 26, Appl
30	919	39.5	590	9	US-09-934-060A-13	Sequence 13, Appl
31	919	39.5	720	9	US-09-934-060A-2	Sequence 2, Appli
32	919	39.5	720	9	US-09-934-060A-4	Sequence 4, Appli
33	911	39.2	178	9	US-09-934-060A-26	Sequence 26, Appl
34	881	37.9	184	14	US-10-024-329-33	Sequence 33, Appl
35	494	21.2	94	11	US-09-891-119A-10	Sequence 10, Appl
36	486	20.9	93	14	US-10-105-545-26	Sequence 26, Appl
37	373	16.0	73	9	US-09-929-924-16	Sequence 16, Appl
38	371	16.0	73	9	US-09-929-924-36	Sequence 36, Appl
39	371	16.0	73	9	US-09-929-924-37	Sequence 37, Appl
40	368	15.8	73	9	US-09-929-924-33	Sequence 33, Appl
41	367	15.8	73	9	US-09-929-924-34	Sequence 34, Appl
42	367	15.8	73	9	US-09-929-924-35	Sequence 35, Appl
43	337	14.5	84	9	US-09-135-238B-8	Sequence 8, Appli
44	293	12.6	240	9	US-09-997-165-2	Sequence 2, Appli
45	288	12.4	58	10	US-09-939-537-35	Sequence 35, Appl
46	238	10.2	95	14	US-10-105-545-25	Sequence 25, Appl
47	197	8.5	72	9	US-09-929-924-17	Sequence 17, Appl
48	162	7.0	570	16	US-10-311-823-13	Sequence 13, Appl
49	162	7.0	586	16	US-10-311-823-7	Sequence 7, Appli
50	157	6.8	1315	15	US-10-094-886-52	Sequence 52, Appl
51	157	6.8	1386	15	US-10-094-886-38	Sequence 38, Appl
52	155	6.7	739	13	US-10-047-542-69	Sequence 69, Appl
53	155	6.7	739	14	US-10-207-655-139	Sequence 139, App
54	155	6.7	739	14	US-10-234-041-7	Sequence 7, Appli
55	152	6.5	1479	12	US-10-231-956A-325	Sequence 325, App
56	152	6.5	1496	12	US-10-211-462-87	Sequence 87, Appl
57	152	6.5	1496	14	US-10-021-660-125	Sequence 125, App
58	152	6.5	1496	15	US-10-331-496A-28	Sequence 28, Appl
59	152	6.5	1498	12	US-10-276-774-1957	Sequence 1957, Ap
60	152	6.5	1498	12	US-10-243-552-899	Sequence 899, App
61	150	6.5	50	14	US-10-076-674-4	Sequence 4, Appli
62	150	6.5	50	14	US-10-076-674-5	Sequence 5, Appli
63	150	6.5	50	15	US-10-355-161A-4	Sequence 4, Appli
64	150	6.5	50	15	US-10-355-161A-5	Sequence 5, Appli
65	146	6.3	1477	14	US-10-274-583-20	Sequence 20, Appl
66	145	6.2	701	15	US-10-168-417A-4	Sequence 4, Appli
67	143	6.2	647	14	US-10-176-847-32	Sequence 32, Appl
68	143	6.2	647	14	US-10-234-041-6	Sequence 6, Appli
69	143	6.2	5636	14	US-10-032-189-128	Sequence 128, App
70	143	6.2	5636	15	US-10-120-801-72	Sequence 72, Appl
71	143	6.2	5636	15	US-10-023-634-93	Sequence 93, Appl
72	143	6.2	5636	16	US-10-408-765A-1895	Sequence 1895, Ap
73	141	6.1	956	15	US-10-108-260A-3037	Sequence 3037, Ap
74	140.5	6.0	737	9	US-09-925-301-1133	Sequence 1133, Ap
75	139.5	6.0	374	16	US-10-311-823-16	Sequence 16, Appl
76	138.5	6.0	442	16	US-10-311-823-12	Sequence 12, Appl
77	138.5	6.0	458	16	US-10-311-823-4	Sequence 4, Appli
78	138	5.9	2828	9	US-09-905-129-21	Sequence 21, Appl
79	138	5.9	2828	9	US-09-991-630-21	Sequence 21, Appl
80	138	5.9	2828	12	US-10-072-012-607	Sequence 607, App
81	138	5.9	2828	12	US-10-072-012-608	Sequence 608, App
82	138	5.9	2828	12	US-10-454-351-21	Sequence 21, Appl
83	138	5.9	2828	12	US-09-802-318-21	Sequence 21, Appl
84	138	5.9	2828	14	US-10-176-847-54	Sequence 54, Appl
85	138	5.9	2828	14	US-10-177-293-110	Sequence 110, App
86	138	5.9	2828	14	US-10-301-822-49	Sequence 49, Appl
87	138	5.9	2828	14	US-10-032-189-126	Sequence 126, App
88	138	5.9	2828	15	US-10-295-027-58	Sequence 58, Appl

89	138	5.9	2828	15	US-10-295-027-1175	Sequence 1175, Ap
90	137.5	5.9	702	12	US-10-253-286-282	Sequence 282, App
91	137.5	5.9	702	12	US-10-380-136-16	Sequence 16, Appl
92	137.5	5.9	702	14	US-10-157-031-341	Sequence 341, App
93	137.5	5.9	702	14	US-10-207-655-87	Sequence 87, Appl
94	137.5	5.9	702	15	US-10-117-937-592	Sequence 592, App
95	137.5	5.9	702	15	US-10-295-027-332	Sequence 332, App
96	137.5	5.9	702	15	US-10-245-871-282	Sequence 282, App
97	137.5	5.9	734	9	US-09-756-551A-17	Sequence 17, Appl
98	137	5.9	5635	16	US-10-451-168-78	Sequence 78, Appl
99	136.5	5.9	595	14	US-10-106-698-4871	Sequence 4871, Ap
100	136.5	5.9	1395	9	US-09-808-602-67	Sequence 67, Appl
101	136.5	5.9	1395	10	US-09-800-198-56	Sequence 56, Appl
102	136.5	5.9	1395	14	US-10-289-776-15	Sequence 15, Appl
103	136	5.8	512	15	US-10-094-749-2493	Sequence 2493, Ap
104	134.5	5.8	1260	13	US-10-024-918-27	Sequence 27, Appl
105	134.5	5.8	1260	16	US-10-650-509-27	Sequence 27, Appl
106	134.5	5.8	3931	15	US-10-120-801-18	Sequence 18, Appl
107	133.5	5.7	6620	15	US-10-080-334-290	Sequence 290, App
108	133.5	5.7	6620	16	US-10-408-765A-2291	Sequence 2291, Ap
109	133.5	5.7	7968	13	US-10-077-130-5	Sequence 5, Appli
110	133	5.7	2783	15	US-10-369-493-6344	Sequence 6344, Ap
111	132.5	5.7	352	12	US-10-042-865-89	Sequence 89, Appl
112	132.5	5.7	352	16	US-10-038-854-156	Sequence 156, App
113	132.5	5.7	1596	16	US-10-408-765A-992	Sequence 992, App
114	132.5	5.7	2328	16	US-10-476-397-4	Sequence 4, Appli
115	132.5	5.7	4675	15	US-10-093-463-74	Sequence 74, Appl
116	132.5	5.7	4691	15	US-10-093-463-72	Sequence 72, Appl
117	132	5.7	1011	12	US-09-825-751A-65	Sequence 65, Appl
118	131	5.6	1257	15	US-10-116-275-202	Sequence 202, App
119	131	5.6	1257	15	US-10-411-010-22	Sequence 22, Appl
120	131	5.6	1257	15	US-10-411-010-23	Sequence 23, Appl
121	131	5.6	5175	15	US-10-120-801-74	Sequence 74, Appl
122	131	5.6	5175	15	US-10-369-493-6859	Sequence 6859, Ap
123	131	5.6	5175	15	US-10-369-493-6861	Sequence 6861, Ap
124	131	5.6	5198	12	US-10-346-863-34	Sequence 34, Appl
125	131	5.6	5198	15	US-10-120-801-75	Sequence 75, Appl

ALIGNMENTS

```
RESULT 1
US-09-939-537-5
; Sequence 5, Application US/09939537
; Publication No. US20030138410A1
;   GENERAL INFORMATION:
;        APPLICANT: Seed, Brian
;                   Banapour, Babak
;                   Romeo, Charles
;                   Kolanus, Waldemar
;        TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
;                            CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
;        NUMBER OF SEQUENCES: 53
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Clark & Elbing LLP
;             STREET: 176 Federal Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02110
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/939,537
;             FILING DATE: 24-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/284,391
;             FILING DATE: 02-AUG-1994
;             APPLICATION NUMBER: 08/195,395
;             FILING DATE: 14-FEB-1994
;             APPLICATION NUMBER: 07/847,566
;             FILING DATE: 06-MAR-1992
;             APPLICATION NUMBER: 07/665,961
;             FILING DATE: 07-MAR-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Elbing, Karen L
;             REGISTRATION NUMBER: 35,238
;             REFERENCE/DOCKET NUMBER: 00786/247001
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617-428-0200
;             TELEFAX: 617-428-7045
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 5:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 462 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-939-537-5

  Query Match            87.7%;  Score 2040;  DB 10;  Length 462;
  Best Local Similarity  91.4%;  Pred. No. 1.5e-149;
  Matches  405;  Conservative    1;  Mismatches    9;  Indels   28;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||||||||                   |||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH---------------ADPQ------ 399

Qy        421 DPPAASALPAALAVISFLLGLGL 443
                    |   |  | || |: |
Db        400 -------LCYILDAILFLYGIVL 415


RESULT 2
US-09-243-008-5
; Sequence 5, Application US/09243008
; Publication No. US20040005334A1
;   GENERAL INFORMATION:
;        APPLICANT: Seed, Brian et al.
;        TITLE OF INVENTION: Redirection of Cellular Immunity by
;                            Receptor Chimeras
;        NUMBER OF SEQUENCES: 40
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Fish & Richardson P.C.
;             STREET: 225 Franklin Street
```

```
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02110-2804
;       COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;             COMPUTER: IBM PS/2 Model 50Z or 55SX
;             OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
;             SOFTWARE: Wordperfect (Version 5.0)
;       CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/243,008
;             FILING DATE: 02-Feb-1999
;       PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/394,176
;             FILING DATE: SEPTEMBER 11,1995
;             APPLICATION NUMBER: 08/203,866
;             FILING DATE: February 28, 1994
;             APPLICATION NUMBER: 07/847,566
;             FILING DATE: March 6, 1992
;             APPLICATION NUMBER: 07/665,961
;             FILING DATE: March 7, 1991
;       ATTORNEY/AGENT INFORMATION:
;             NAME: Karen F. Lech, Ph.D
;             REGISTRATION NUMBER: 35,238
;             REFERENCE/DOCKET NUMBER: 00786/270001
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617) 542-5070
;             TELEFAX: (617) 542-8906
;             TELEX: 200154
;   INFORMATION FOR SEQ ID NO: 5:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 462 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-243-008-5

  Query Match             87.7%;  Score 2040;  DB 11;  Length 462;
  Best Local Similarity   91.4%;  Pred. No. 1.5e-149;
  Matches  405;  Conservative    1;  Mismatches    9;  Indels   28;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH----------------ADPQ------ 399

Qy        421 DPPAASALPAALAVISFLLGLGL 443
                  |   |  | || |: |
Db        400 -------LCYILDAILFLYGIVL 415
```

```
RESULT 3
US-09-939-537-4
; Sequence 4, Application US/09939537
; Publication No. US20030138410A1
;   GENERAL INFORMATION:
;       APPLICANT: Seed, Brian
;                  Banapour, Babak
;                  Romeo, Charles
;                  Kolanus, Waldemar
;       TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
;                           CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
;       NUMBER OF SEQUENCES: 53
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Clark & Elbing LLP
;            STREET: 176 Federal Street
;            CITY: Boston
;            STATE: MA
;            COUNTRY: USA
;            ZIP: 02110
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/939,537
;            FILING DATE: 24-Aug-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/284,391
;            FILING DATE: 02-AUG-1994
;            APPLICATION NUMBER: 08/195,395
;            FILING DATE: 14-FEB-1994
;            APPLICATION NUMBER: 07/847,566
;            FILING DATE: 06-MAR-1992
;            APPLICATION NUMBER: 07/665,961
;            FILING DATE: 07-MAR-1991
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Elbing, Karen L
;            REGISTRATION NUMBER: 35,238
;            REFERENCE/DOCKET NUMBER: 00786/247001
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 617-428-0200
;            TELEFAX: 617-428-7045
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 575 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-939-537-4

  Query Match             87.7%;  Score 2038.5;  DB 10;  Length 575;
  Best Local Similarity   92.1%;  Pred. No. 2.6e-149;
  Matches  406;  Conservative    4;  Mismatches    8;  Indels   23;  Gaps    3;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
```

```
Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PR-----------------ASAL 401
              ||||||||||||||||||||||||||||||||||  |:                   | |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPKLCYLLDGILFIYGVIITALYL 420

Qy        402 PAPPTGSALPDPQTASALPDP 422
              |  : ||    :||: | ||
Db        421 RAKFSRSA----ETAANLQDP 437


RESULT 4
US-09-243-008-4
; Sequence 4, Application US/09243008
; Publication No. US20040005334A1
;   GENERAL INFORMATION:
;        APPLICANT: Seed, Brian et al.
;        TITLE OF INVENTION: Redirection of Cellular Immunity by
;                            Receptor Chimeras
;        NUMBER OF SEQUENCES: 40
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Fish & Richardson P.C.
;             STREET: 225 Franklin Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02110-2804
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;             COMPUTER: IBM PS/2 Model 50Z or 55SX
;             OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
;             SOFTWARE: Wordperfect (Version 5.0)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/243,008
;             FILING DATE: 02-Feb-1999
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/394,176
;             FILING DATE: SEPTEMBER 11,1995
;             APPLICATION NUMBER: 08/203,866
;             FILING DATE: February 28, 1994
;             APPLICATION NUMBER: 07/847,566
;             FILING DATE: March 6, 1992
;             APPLICATION NUMBER: 07/665,961
;             FILING DATE: March 7, 1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Karen F. Lech, Ph.D
;             REGISTRATION NUMBER: 35,238
;             REFERENCE/DOCKET NUMBER: 00786/270001
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (617) 542-5070
;             TELEFAX: (617) 542-8906
;             TELEX: 200154
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 575 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-243-008-4

  Query Match             87.7%;  Score 2038.5;  DB 11;  Length 575;
  Best Local Similarity   92.1%;  Pred. No. 2.6e-149;
  Matches  406;  Conservative    4;  Mismatches    8;  Indels   23;  Gaps    3;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH--PR-----------------ASAL 401
              ||||||||||||||||||||||||||||||||||  |:                   | |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHADPKLCYLLDGILFIYGVIITALYL 420

Qy        402 PAPPTGSALPDPQTASALPDP 422
              |  : ||    :||: | ||
Db        421 RAKFSRSA----ETAANLQDP 437


RESULT 5
US-09-891-119A-9
; Sequence 9, Application US/09891119A
; Publication No. US20040013683A1
; GENERAL INFORMATION:
;  APPLICANT: Maddon, Paul J.
;  TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4
;  FILE REFERENCE: 24577-CY-B
;  CURRENT APPLICATION NUMBER: US/09/891,119A
;  CURRENT FILING DATE:  2001-06-25
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
;   LENGTH: 457
;   TYPE: PRT
;   ORGANISM: human
US-09-891-119A-9

  Query Match             87.7%;  Score 2038;  DB 11;  Length 457;
  Best Local Similarity   89.4%;  Pred. No. 2.1e-149;
  Matches  406;  Conservative    2;  Mismatches   16;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
```

```
Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                  ||      |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 6
US-09-939-537-29
; Sequence 29, Application US/09939537
; Publication No. US20030138410A1
;  GENERAL INFORMATION:
;       APPLICANT: Seed, Brian
;                  Banapour, Babak
;                  Romeo, Charles
;                  Kolanus, Waldemar
;       TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
;                           CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
;       NUMBER OF SEQUENCES: 53
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Clark & Elbing LLP
;            STREET: 176 Federal Street
;            CITY: Boston
;            STATE: MA
;            COUNTRY: USA
;            ZIP: 02110
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/939,537
;            FILING DATE: 24-Aug-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/284,391
;            FILING DATE: 02-AUG-1994
;            APPLICATION NUMBER: 08/195,395
;            FILING DATE: 14-FEB-1994
;            APPLICATION NUMBER: 07/847,566
;            FILING DATE: 06-MAR-1992
;            APPLICATION NUMBER: 07/665,961
;            FILING DATE: 07-MAR-1991
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Elbing, Karen L
;            REGISTRATION NUMBER: 35,238
;            REFERENCE/DOCKET NUMBER: 00786/247001
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 617-428-0200
;            TELEFAX: 617-428-7045
;            TELEX: <Unknown>
;  INFORMATION FOR SEQ ID NO: 29:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 398 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-09-939-537-29

  Query Match             87.6%;  Score 2037;  DB 10;  Length 398;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-149;
  Matches  395;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 7
US-09-939-537-6
; Sequence 6, Application US/09939537
; Publication No. US20030138410A1
;  GENERAL INFORMATION:
;       APPLICANT: Seed, Brian
;                  Banapour, Babak
;                  Romeo, Charles
;                  Kolanus, Waldemar
;       TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
;                           CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
;       NUMBER OF SEQUENCES: 53
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Clark & Elbing LLP
;            STREET: 176 Federal Street
;            CITY: Boston
;            STATE: MA
;            COUNTRY: USA
;            ZIP: 02110
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: FastSEQ for Windows Version 2.0
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/939,537
;            FILING DATE: 24-Aug-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/284,391
;            FILING DATE: 02-AUG-1994
;            APPLICATION NUMBER: 08/195,395
;            FILING DATE: 14-FEB-1994
;            APPLICATION NUMBER: 07/847,566
;            FILING DATE: 06-MAR-1992
;            APPLICATION NUMBER: 07/665,961
;            FILING DATE: 07-MAR-1991
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Elbing, Karen L
;            REGISTRATION NUMBER: 35,238
;            REFERENCE/DOCKET NUMBER: 00786/247001
```

```
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 617-428-0200
;            TELEFAX: 617-428-7045
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 532 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-939-537-6

  Query Match            87.6%;  Score 2037;  DB 10;  Length 532;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-149;
  Matches  395;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 8
US-09-243-008-6
; Sequence 6, Application US/09243008
; Publication No. US20040005334A1
;   GENERAL INFORMATION:
;       APPLICANT: Seed, Brian et al.
;       TITLE OF INVENTION: Redirection of Cellular Immunity by
;                           Receptor Chimeras
;       NUMBER OF SEQUENCES: 40
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fish & Richardson P.C.
;            STREET: 225 Franklin Street
;            CITY: Boston
;            STATE: MA
;            COUNTRY: USA
;            ZIP: 02110-2804
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;            COMPUTER: IBM PS/2 Model 50Z or 55SX
;            OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
;            SOFTWARE: Wordperfect (Version 5.0)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/243,008
;            FILING DATE: 02-Feb-1999
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/394,176
;            FILING DATE: SEPTEMBER 11,1995
;            APPLICATION NUMBER: 08/203,866
;            FILING DATE: February 28, 1994
;            APPLICATION NUMBER: 07/847,566
;            FILING DATE: March 6, 1992
;            APPLICATION NUMBER: 07/665,961
;            FILING DATE: March 7, 1991
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Karen F. Lech, Ph.D
;            REGISTRATION NUMBER: 35,238
;            REFERENCE/DOCKET NUMBER: 00786/270001
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (617) 542-5070
;            TELEFAX: (617) 542-8906
;            TELEX: 200154
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 532 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-243-008-6

  Query Match            87.6%;  Score 2037;  DB 11;  Length 532;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-149;
  Matches  395;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395
              |||||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVH 395


RESULT 9
US-10-103-597A-39
; Sequence 39, Application US/10103597A
; Publication No. US20030096432A1
; GENERAL INFORMATION:
;  APPLICANT: Jakobsen, Bent Karsten
;  TITLE OF INVENTION: Screening Methods
;  FILE REFERENCE: 102286.142
;  CURRENT APPLICATION NUMBER: US/10/103,597A
;  CURRENT FILING DATE:  2002-10-17
;  PRIOR APPLICATION NUMBER: PCT/GB00/03579
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: GB 9922352.1
;  PRIOR FILING DATE: 1999-09-21
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 39
;   LENGTH: 458
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-103-597A-39

  Query Match             87.4%;  Score 2032;  DB 14;  Length 458;
  Best Local Similarity   89.2%;  Pred. No. 6.1e-149;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |    || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 10
US-10-188-444-39
; Sequence 39, Application US/10188444
; Publication No. US20030104635A1
; GENERAL INFORMATION:
;  APPLICANT: Jakobsen, Bent Karsten
;  TITLE OF INVENTION: Screening Methods
;  FILE REFERENCE: 102286.142 (CIP)
;  CURRENT APPLICATION NUMBER: US/10/188,444
;  CURRENT FILING DATE:  2002-07-02
;  PRIOR APPLICATION NUMBER: PCT/GB00/03579
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: GB 9922352.1
;  PRIOR FILING DATE: 1999-09-21
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 39
;   LENGTH: 458
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-188-444-39

  Query Match             87.4%;  Score 2032;  DB 14;  Length 458;
  Best Local Similarity   89.2%;  Pred. No. 6.1e-149;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |    || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |    || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 11
US-10-207-655-170
; Sequence 170, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
;  APPLICANT: Ledbetter, Jeffrey A.
;  APPLICANT:  Hayden-Ledbetter, Martha S.
;  TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
;  FILE REFERENCE: 390069.401C1
;  CURRENT APPLICATION NUMBER: US/10/207,655
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 426
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 170
;   LENGTH: 458
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-207-655-170

  Query Match             87.4%;  Score 2032;  DB 14;  Length 458;
  Best Local Similarity   89.2%;  Pred. No. 6.1e-149;
  Matches  405;  Conservative    2;  Mismatches   17;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 12
US-10-151-274-3
; Sequence 3, Application US/10151274
; Publication No. US20030064071A1
; GENERAL INFORMATION:
;  APPLICANT: Littman, Dan R.
;  APPLICANT:  Kwon, Douglas S.
;  APPLICANT:  van Kooyk, Yvette
;  APPLICANT:  Geijtenbeck, Tneo
;  TITLE OF INVENTION: METHODS OF USING A FACILITATOR OF RETROVIRAL ENTRY
;  TITLE OF INVENTION:  INTO
;  TITLE OF INVENTION:  CELLS
;  FILE REFERENCE: 1049-1-017
;  CURRENT APPLICATION NUMBER: US/10/151,274
;  CURRENT FILING DATE:  2002-05-20
;  PRIOR APPLICATION NUMBER: US/09/517,605
;  PRIOR FILING DATE: 2000-03-02
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 458
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-151-274-3

  Query Match             87.1%;  Score 2024;  DB 12;  Length 458;
  Best Local Similarity   89.0%;  Pred. No. 2.6e-148;
  Matches  404;  Conservative    2;  Mismatches   18;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKWVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 13
US-10-097-044A-1
; Sequence 1, Application US/10097044A
; Publication No. US20030143220A1
;  GENERAL INFORMATION:
;       APPLICANT: Capon, Daniel J.
;                  Gregory, Timothy J.
;       TITLE OF INVENTION: Adheson Variants
;       NUMBER OF SEQUENCES: 25
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 460 Point San Bruno Blvd
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: patin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/097,044A
;            FILING DATE: 28-May-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/457,918
;            FILING DATE: 1-JUN-1995
;            APPLICATION NUMBER: 08/236311
;            FILING DATE: 02-MAY-1994
;            APPLICATION NUMBER: 07/936190
;            FILING DATE: 26-AUG-1992
;            APPLICATION NUMBER: 07/842777
;            FILING DATE: 18-FEB-1992
;            APPLICATION NUMBER: 07/250785
;            FILING DATE: 28-SEP-1988
;            APPLICATION NUMBER: 07/104329
;            FILING DATE: 02-OCT-1987
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Kubinec, Jeffrey S.
;            REGISTRATION NUMBER: 36,575
;            REFERENCE/DOCKET NUMBER: P0444P1C3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 415/225-8228
;            TELEFAX: 415/952-9881
;            TELEX: 910/371-7168
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 402 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-097-044A-1

  Query Match             86.8%;  Score 2017;  DB 14;  Length 402;
  Best Local Similarity   99.7%;  Pred. No. 7.4e-148;
  Matches  392;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTP 393
              |||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTP 393


RESULT 14
US-08-681-219-27
; Sequence 27, Application US/08681219
; Publication No. US20020058607A1
;  GENERAL INFORMATION:
;    APPLICANT:  Takaaki Sato and Junn Yanagisawa
;    TITLE OF INVENTION:  COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN
;    TITLE OF INVENTION:  SIGNAL-TRANSDUCING PROTEINS AND THE GLGF
;    TITLE OF INVENTION:  (PDZ/DHR) DOMAIN AND USES THEREOF
;    NUMBER OF SEQUENCES:  35
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/681,219
;      FILING DATE:  22-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0575/48962/JPW/JKM
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  458 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-681-219-27

  Query Match           86.6%;  Score 2014;  DB 8;  Length 458;
  Best Local Similarity   88.5%;  Pred. No. 1.5e-147;
  Matches  402;  Conservative    3;  Mismatches   19;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTITLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||  |||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQENVLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||      |    || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 15
US-09-230-111C-25
; Sequence 25, Application US/09230111C
; Publication No. US20030203414A1
; GENERAL INFORMATION:
;  APPLICANT: Sato, Taka-Aki
;  APPLICANT:  Yanagisawa, Junn
;  TITLE OF INVENTION: COMPOUNDS THAT INHIBIT INTERACTION BETWEEN
;  TITLE OF INVENTION:  SIGNAL-TRANSDUCING PROTEINS AND THE GLGF (PDZ/DHR)
;  TITLE OF INVENTION:  DOMAIN AND USES THEREOF
;  FILE REFERENCE: 48962-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/230,111C
;  CURRENT FILING DATE:  1999-05-17
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
;   LENGTH: 458
;   TYPE: PRT
;   ORGANISM: human
US-09-230-111C-25

  Query Match           86.6%;  Score 2014;  DB 11;  Length 458;
  Best Local Similarity   88.5%;  Pred. No. 1.5e-147;
  Matches  402;  Conservative    3;  Mismatches   19;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              |||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTITLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||  |||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQENVLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396
```

```
Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                  ||      |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 16
US-10-092-138-25
; Sequence 25, Application US/10092138
; Publication No. US20030170723A1
; GENERAL INFORMATION:
;  APPLICANT: Sato, Taka-Aki
;  TITLE OF INVENTION: METHOD OF PREPARING A PROTEIN ARRAY BASED ON
;  TITLE OF INVENTION:  BIOCHEMICAL PROTEIN-PROTEIN INTERACTION
;  FILE REFERENCE: 65823/JPW/PT
;  CURRENT APPLICATION NUMBER: US/10/092,138
;  CURRENT FILING DATE:  2002-09-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
;   LENGTH: 458
;   TYPE: PRT
;   ORGANISM: human
US-10-092-138-25

  Query Match             86.6%;  Score 2014;  DB 14;  Length 458;
  Best Local Similarity   88.5%;  Pred. No. 1.5e-147;
  Matches  402;  Conservative    3;  Mismatches   19;  Indels   30;  Gaps    2;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTITLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||  |||||||||||||||||||||||||||||||||||||||||| |||
Db        301 LEAKTGKLHQENVLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||| |
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 396

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                  ||      |   || :|||:  ||  | |
Db        397 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 425


RESULT 17
US-09-891-119A-2
; Sequence 2, Application US/09891119A
; Publication No. US20040013683A1
; GENERAL INFORMATION:
;  APPLICANT: Maddon, Paul J.
;  TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4
;  FILE REFERENCE: 24577-CY-B
;  CURRENT APPLICATION NUMBER: US/09/891,119A
;  CURRENT FILING DATE:  2001-06-25
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 397
;   TYPE: PRT
;   ORGANISM: Human
US-09-891-119A-2

  Query Match             86.1%;  Score 2001;  DB 11;  Length 397;
  Best Local Similarity   98.7%;  Pred. No. 1.3e-146;
  Matches  389;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||| |||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db         61 ILGNQGSSLTKGPSKLNDRADSRRSLWDQGNFPLIIRNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVDFSFPLAFTVEKLTGSGELWW 240

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 360

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394
              |||||||||||||||||||||||||||||||||
Db        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPV 394


RESULT 18
US-10-097-044A-4
; Sequence 4, Application US/10097044A
; Publication No. US20030143220A1
;  GENERAL INFORMATION:
;       APPLICANT: Capon, Daniel J.
;                  Gregory, Timothy J.
;       TITLE OF INVENTION: Adheson Variants
;       NUMBER OF SEQUENCES: 25
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 460 Point San Bruno Blvd
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: patin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/097,044A
;            FILING DATE: 28-May-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/457,918
;            FILING DATE: 1-JUN-1995
;            APPLICATION NUMBER: 08/236311
;            FILING DATE: 02-MAY-1994
;            APPLICATION NUMBER: 07/936190
;            FILING DATE: 26-AUG-1992
;            APPLICATION NUMBER: 07/842777
```

```
;             FILING DATE: 18-FEB-1992
;             APPLICATION NUMBER: 07/250785
;             FILING DATE: 28-SEP-1988
;             APPLICATION NUMBER: 07/104329
;             FILING DATE: 02-OCT-1987
;       ATTORNEY/AGENT INFORMATION:
;             NAME: Kubinec, Jeffrey S.
;             REGISTRATION NUMBER: 36,575
;             REFERENCE/DOCKET NUMBER: P0444P1C3
;       TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415/225-8228
;             TELEFAX: 415/952-9881
;             TELEX: 910/371-7168
;   INFORMATION FOR SEQ ID NO: 4:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 434 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-097-044A-4

  Query Match            81.9%;  Score 1904;  DB 14;  Length 434;
  Best Local Similarity   99.7%;  Pred. No. 4.8e-139;
  Matches  369;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         56 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 115

Qy         84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTL 175

Qy        144 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 ESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 235

Qy        204 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 FQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSV 295

Qy        264 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 KRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQ 355

Qy        324 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESN 383
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        356 KNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESN 415

Qy        384 IKVLPTWSTP 393
              ||||||||||
Db        416 IKVLPTWSTP 425


RESULT 19
US-09-759-841-6
; Sequence 6, Application US/09759841
; Patent No. US20010039026A1
; GENERAL INFORMATION:
;  APPLICANT:  Rickett, Graham A
;  APPLICANT:  Dobbs, Susan
;  APPLICANT:  Perros, Manoussos
;  TITLE OF INVENTION:  Assay Method
;  FILE REFERENCE:  PC10348APME
;  CURRENT APPLICATION NUMBER: US/09/759,841
;  CURRENT FILING DATE:  2001-01-12
;  PRIOR APPLICATION NUMBER:  GB 0000661.9
;  PRIOR FILING DATE:  2000-01-12
;  PRIOR APPLICATION NUMBER:  GB 0000663.5
;  PRIOR FILING DATE:  2000-01-12
;  PRIOR APPLICATION NUMBER:  GB 0000659.3
;  PRIOR FILING DATE:  2000-01-12
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 6
;   LENGTH: 370
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-759-841-6

  Query Match            81.7%;  Score 1899;  DB 9;  Length 370;
  Best Local Similarity   99.7%;  Pred. No. 9.3e-139;
  Matches  368;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 181

Qy        207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITFDLKNKEVSVKRV 241

Qy        267 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNL 301

Qy        327 TCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIKV 386
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        302 TCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNIKV 361

Qy        387 LPTWSTPVH 395
              |||||||||
Db        362 LPTWSTPVH 370


RESULT 20
US-10-024-329-32
; Sequence 32, Application US/10024329
; Publication No. US20030157063A1
; GENERAL INFORMATION:
;  APPLICANT: SANHADJI, Kamel
;  APPLICANT:  TOURAINE, Jean-Louis
;  APPLICANT:  LEROY, Pierre
;  APPLICANT:  MEHTALI, Majid
;  TITLE OF INVENTION: Gene therapy using anti-gp41 antibody and cd4 immunoadhesin
;  FILE REFERENCE: 109993
;  CURRENT APPLICATION NUMBER: US/10/024,329
;  CURRENT FILING DATE:  2001-12-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
;   LENGTH: 448
;   TYPE: PRT
;   ORGANISM: human sCD4
US-10-024-329-32

  Query Match            81.5%;  Score 1894.5;  DB 14;  Length 448;
  Best Local Similarity   87.0%;  Pred. No. 2.7e-138;
  Matches  395;  Conservative    2;  Mismatches   18;  Indels   39;  Gaps   10;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||| |||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPF-HLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 59

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              |||||||||||||||||||||||||||||||||||||||||||||||||    ||||||||
Db         60 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC--VDQKEEVQL 117
```

```
Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              |||||||||||||||||||||||||||||||||||||||||| ||| |||||||| |||||||
Db        118 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPR-KNI-GGKTLSVS-LELQDSG 174

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWW 240
              |||||||| |||||||||||||| |||||||||||||||||||||||||||||| ||||
Db        175 TWTCTVLQN-KKVEFKIDIVVLAF-KASSIVYKKEGEQVEFSFPLAFTVEKLTGS-ELWW 231

Qy        241 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 QAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLA 291

Qy        301 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWV 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        292 LEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWV 351

Qy        361 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
              |||||||||||||||||||||||||||||||| |
Db        352 LNPEAGMWQCLLSDSGQVLLESNIKVLPTWSTPVQP---------------------- 387

Qy        421 DPPAASALPAALAVISFLLGLGLGV-ACVLARTR 453
                    ||     |   || :|||:  ||  | |
Db        388 -----MALIVLGGVAGLLLFIGLGIFFCVRCRHR 416


RESULT 21
US-08-485-163-7
; Sequence 7, Application US/08485163
; Publication No. US20020098191A1
;  GENERAL INFORMATION:
;    APPLICANT:  Beaudry, Gary A.
;    APPLICANT:  Maddon, Paul J.
;    TITLE OF INVENTION:  CD4-GAMMA2 CD4-IgG2 CHIMERAS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,163
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  37690-II-1-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  310 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  homo sapien
;      CELL TYPE:  lymphocyte
US-08-485-163-7

  Query Match             45.2%;  Score 1050.5;  DB 8;  Length 310;
  Best Local Similarity   75.4%;  Pred. No. 4.4e-73;
  Matches  221;  Conservative   12;  Mismatches   41;  Indels   19;  Gaps    4;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKA--SSIVYKKEGEQVEFSFPLAFTVEKLTG---- 234
              |||||||||||||||||||||||| |  |  ::    ||::     |  |  |
Db        181 TWTCTVLQNKKVEFKIDIVVLAFTVAAPSVFIFPPSDEQLKSG--TASVVCLLNNFYPR 238

Qy        235 SGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQA 287
                :: |: : |  | :           | : ||:          |   ||| :|
Db        239 EAKVQWKVDNALQSGN----------SQESVTEQDSKDSTYSLSSTLTLSKA 280


RESULT 22
US-09-766-995-6
; Sequence 6, Application US/09766995
; Patent No. US20020052481A1
; GENERAL INFORMATION:
;  APPLICANT: Graham P. Allaway et al.
;  TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED CD4-GAMMA2 AND CD4-IgG2 IMMUNOCONJU
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 2048/41215-CB/JPW/SHS
;  CURRENT APPLICATION NUMBER: US/09/766,995
;  CURRENT FILING DATE:  2001-01-22
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
;   LENGTH: 310
;   TYPE: PRT
;   ORGANISM: homo sapians
US-09-766-995-6

  Query Match             45.2%;  Score 1050.5;  DB 9;  Length 310;
  Best Local Similarity   75.4%;  Pred. No. 4.4e-73;
  Matches  221;  Conservative   12;  Mismatches   41;  Indels   19;  Gaps    4;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKA--SSIVYKKEGEQVEFSFPLAFTVEKLTG---- 234
              |||||||||||||||||||||||| |  |  ::    ||::     |  |  |
Db        181 TWTCTVLQNKKVEFKIDIVVLAFTVAAPSVFIFPPSDEQLKSG--TASVVCLLNNFYPR 238

Qy        235 SGELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQA 287
                :: |: : |  | :           | : ||:          |   ||| :|
Db        239 EAKVQWKVDNALQSGN----------SQESVTEQDSKDSTYSLSSTLTLSKA 280


RESULT 23
US-08-485-163-5
; Sequence 5, Application US/08485163
; Publication No. US20020098191A1
;  GENERAL INFORMATION:
;    APPLICANT:  Beaudry, Gary A.
;    APPLICANT:  Maddon, Paul J.
```

```
;    TITLE OF INVENTION:  CD4-GAMMA2 CD4-IgG2 CHIMERAS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,163
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  37690-II-1-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  530 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    ORIGINAL SOURCE:
;      ORGANISM:  homo sapien
;      CELL TYPE:  lymphocyte
US-08-485-163-5

  Query Match             45.0%;  Score 1046;  DB 8;  Length 530;
  Best Local Similarity   93.4%;  Pred. No. 2e-72;
  Matches  211;  Conservative    2;  Mismatches    3;  Indels   10;  Gaps    3;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLA 226
              |||||||||||||||||||||  ||:     :| |   ||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAF--AST-----KGPSV---FPLA 216


RESULT 24
US-09-766-995-4
; Sequence 4, Application US/09766995
; Patent No. US20020052481A1
; GENERAL INFORMATION:
;  APPLICANT: Graham P. Allaway et al.
;  TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED CD4-GAMMA2 AND CD4-IgG2 IMMUNOCONJ
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 2048/41215-CB/JPW/SHS
;  CURRENT APPLICATION NUMBER: US/09/766,995
;  CURRENT FILING DATE:  2001-01-22
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
;   LENGTH: 530
;   TYPE: PRT
;   ORGANISM: homo sapians
US-09-766-995-4

  Query Match             45.0%;  Score 1046;  DB 9;  Length 530;
  Best Local Similarity   93.4%;  Pred. No. 2e-72;
  Matches  211;  Conservative    2;  Mismatches    3;  Indels   10;  Gaps    3;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLA 226
              |||||||||||||||||||||  ||:     :| |   ||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLAF--AST-----KGPSV---FPLA 216


RESULT 25
US-08-485-163-3
; Sequence 3, Application US/08485163
; Publication No. US20020098191A1
; GENERAL INFORMATION:
;    APPLICANT:  Beaudry, Gary A.
;    APPLICANT:  Maddon, Paul J.
;    TITLE OF INVENTION:  CD4-GAMMA2 CD4-IgG2 CHIMERAS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/485,163
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  37690-II-1-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  432 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  homo sapien
;      CELL TYPE:  lymphocyte
US-08-485-163-3

  Query Match             44.9%;  Score 1045;  DB 8;  Length 432;
  Best Local Similarity   98.5%;  Pred. No. 1.8e-72;
```

```
  Matches  203;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||::
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFER 206


RESULT 26
US-09-766-995-2
; Sequence 2, Application US/09766995
; Patent No. US20020052481A1
; GENERAL INFORMATION:
;  APPLICANT: Graham P. Allaway et al.
;  TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED CD4-GAMMA2 AND CD4-IgG2 IMMUNOCONJ
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 2048/41215-CB/JPW/SHS
;  CURRENT APPLICATION NUMBER: US/09/766,995
;  CURRENT FILING DATE:  2001-01-22
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 432
;   TYPE: PRT
;   ORGANISM: homo sapians
US-09-766-995-2

  Query Match           44.9%;  Score 1045;  DB 9;  Length 432;
  Best Local Similarity  98.5%;  Pred. No. 1.8e-72;
  Matches  203;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||::
Db        181 TWTCTVLQNQKKVEFKIDIVVLAFER 206


RESULT 27
US-09-939-537-31
; Sequence 31, Application US/09939537
; Publication No. US20030138410A1
;   GENERAL INFORMATION:
;        APPLICANT: Seed, Brian
;                   Banapour, Babak
;                   Romeo, Charles
;                   Kolanus, Waldemar
;        TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
;                            CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
;        NUMBER OF SEQUENCES: 53
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Clark & Elbing LLP
;             STREET: 176 Federal Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02110
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/939,537
;             FILING DATE: 24-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/284,391
;             FILING DATE: 02-AUG-1994
;             APPLICATION NUMBER: 08/195,395
;             FILING DATE: 14-FEB-1994
;             APPLICATION NUMBER: 07/847,566
;             FILING DATE: 06-MAR-1992
;             APPLICATION NUMBER: 07/665,961
;             FILING DATE: 07-MAR-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Elbing, Karen L
;             REGISTRATION NUMBER: 35,238
;             REFERENCE/DOCKET NUMBER: 00786/247001
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617-428-0200
;             TELEFAX: 617-428-7045
;             TELEX: <Unknown>
;    INFORMATION FOR SEQ ID NO: 31:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 203 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-939-537-31

  Query Match           44.8%;  Score 1041;  DB 10;  Length 203;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-72;
  Matches  203;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60

Qy         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120

Qy        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180

Qy        181 TWTCTVLQNQKKVEFKIDIVVLA 203
              |||||||||||||||||||||||
Db        181 TWTCTVLQNQKKVEFKIDIVVLA 203


RESULT 28
US-10-125-692-10
; Sequence 10, Application US/10125692
; Publication No. US20030044429A1
; GENERAL INFORMATION:
;  APPLICANT: Aderem, Alan
;  APPLICANT:  Hayashi, Fumitaka
;  APPLICANT:  Smith, Kelly D.
;  APPLICANT:  Underhill, David M.
;  APPLICANT:  Ozinsky, Adrian
;  TITLE OF INVENTION: Toll-Like Receptor 5 Ligands and Methods
```

```
;  TITLE OF INVENTION:  of Use
;  FILE REFERENCE: P-IS 5155
;  CURRENT APPLICATION NUMBER: US/10/125,692
;  CURRENT FILING DATE:  2002-04-17
;  PRIOR APPLICATION NUMBER: US 60/285,477
;  PRIOR FILING DATE: 2001-04-20
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 612
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-125-692-10

  Query Match             42.7%;  Score 993;  DB 14;  Length 612;
  Best Local Similarity   53.8%;  Pred. No. 3.2e-68;
  Matches  211;  Conservative   62;  Mismatches  109;  Indels   10;  Gaps    7;

Qy          1 MNRGVPFRH-LLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQI 59
              | | :  |  |||:|||: | | |||  :||||:|:: || | :|||    | || |:|
Db          1 MCRAISLRRLLLLLLQLSQLLAVTQGKTLVLGKEGESAELPCESSQKKITVFTWKFSDQR 60

Qy         60 KILGNQG-SFLTKG--PSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKE 116
              ||||  |   | :|  ||:  || ||::  |::|:|||||  ||:||| |||||:|::||
Db         61 KILGQHGKGVLIRGGSPSQF-DRFDSKKGAWEKGSFPLIINKLKMEDSQTYICELENRKE 119

Qy        117 EVQLLVFGLTANSDTHLLQGQSLTLTLES-PPGSSPSVQCRSPRGKNIQGGKTLSVSQLE 175
              ||:| || :| :  | |||||||||||:|    |:|  :|:  :|| : | | ||:| |
Db        120 EVELWVFKVTFSPGTSLLQGQSLTLTLDSNSKVSNPLTECKHKKGKVVSGSKVLSMSNLR 179

Qy        176 LQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGS 235
              :|||  | |||  :|||  | : : || ||  :   || |||  |||||| |  |   |
Db        180 VQDSDFWNCTVTLDQKKNWFGMTLSVLGFQSTAITAYKSEGESAEFSFPLNFAEE--NGW 237

Qy        236 GELWWQAERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSG 295
              ||| |:||:  |   : ||:| :|||||||:: |:| |||: : ||| | :||    |:||||
Db        238 GELMWKAEKDSFFQPWISFSIKNKEVSVQKSTKDLKLQLKETLPLTLKIPQVSLQFAGSG 297

Qy        296 NLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKRE 355
              |||| |:   | ||||||||||:  ||   ||||| ||||||: |:|| ||:||:||: :
Db        298 NLTLTLD--KGTLHQEVNLVVMKVAQLNNTLTCEVMGPTSPKMRLTLKQENQEARVSEEQ 355

Qy        356 KPVWVLNPEAGMWQCLLSDSGQVLLESNIKVL 387
              | | |: || |:||||||:  :| ::| |:||
Db        356 KVVQVVAPETGLWQCLLSEGDKVKMDSRIQVL 387


RESULT 29
US-10-073-118-26
; Sequence 26, Application US/10073118
; Publication No. US20030054554A1
; GENERAL INFORMATION:
;  APPLICANT: BECQUART, JEROME
;  TITLE OF INVENTION: ALBUMIN DERIVATIVES WITH THERAPEUTIC FUNCTIONS
;  FILE REFERENCE: 06832.1429-03
;  CURRENT APPLICATION NUMBER: US/10/073,118
;  CURRENT FILING DATE:  2002-02-12
;  PRIOR APPLICATION NUMBER: 09/551,635
;  PRIOR FILING DATE: 2000-04-18
;  PRIOR APPLICATION NUMBER: 09/004,319
;  PRIOR FILING DATE: 1998-01-08
;  PRIOR APPLICATION NUMBER: 08/479,146
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: 08/295,078
;  PRIOR FILING DATE: 1994-08-26
;  PRIOR APPLICATION NUMBER: 08/121,236
;  PRIOR FILING DATE: 1993-09-13
;  PRIOR APPLICATION NUMBER: 07/955,243
;  PRIOR FILING DATE: 1992-10-01
;  PRIOR APPLICATION NUMBER: 07/561,879
;  PRIOR FILING DATE: 1990-08-02
;  PRIOR APPLICATION NUMBER: FR 89 10480
;  PRIOR FILING DATE: 1989-08-03
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
;   LENGTH: 788
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Protein fusion
;   OTHER INFORMATION: prepro-HSA-V1V2
US-10-073-118-26

  Query Match             39.6%;  Score 921;  DB 14;  Length 788;
  Best Local Similarity   95.8%;  Pred. No. 1.7e-62;
  Matches  182;  Conservative    1;  Mismatches    3;  Indels    4;  Gaps    1;

Qy         19 LPAATQG----NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPS 74
              | ||:|      ||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 LVAASQAALGLKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPS 658

Qy         75 KLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLL 134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 KLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLL 718

Qy        135 QGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVE 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 QGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVE 778

Qy        195 FKIDIVVLAF 204
              ||||||||||
Db        779 FKIDIVVLAF 788


RESULT 30
US-09-934-060A-13
; Sequence 13, Application US/09934060A
; Patent No. US20020155121A1
; GENERAL INFORMATION:
;  APPLICANT: DeVico, Anthony L.
;  APPLICANT:  Fouts, Timothy R.
;  APPLICANT:  Tuskan, Robert G.
;  TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE
;  FILE REFERENCE: 4115-144 CIP
;  CURRENT APPLICATION NUMBER: US/09/934,060A
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: US 09/684,026
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: US 60/158,321
;  PRIOR FILING DATE: 1999-10-08
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
;   LENGTH: 590
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthesized construct
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (586)..(586)
;   OTHER INFORMATION: Xaa can be any amino acid
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (589)..(589)
;   OTHER INFORMATION: Xaa can be any amino acid
US-09-934-060A-13

  Query Match             39.5%;  Score 919;  DB 9;  Length 590;
  Best Local Similarity   98.4%;  Pred. No. 1.7e-62;
  Matches  180;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        399 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 458

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        459 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 518

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        519 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAEQK 578

Qy        207 ASS 209
                |
Db        579 LIS 581


RESULT 31
US-09-934-060A-2
; Sequence 2, Application US/09934060A
; Patent No. US20020155121A1
; GENERAL INFORMATION:
;  APPLICANT: DeVico, Anthony L.
;  APPLICANT:  Fouts, Timothy R.
;  APPLICANT:  Tuskan, Robert G.
;  TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE
;  FILE REFERENCE: 4115-144 CIP
;  CURRENT APPLICATION NUMBER: US/09/934,060A
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: US 09/684,026
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: US 60/158,321
;  PRIOR FILING DATE: 1999-10-08
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthesized construct
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (716)..(716)
;   OTHER INFORMATION: Xaa can be any amino acid
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (719)..(719)
;   OTHER INFORMATION: Xaa can be any amino acid
US-09-934-060A-2

  Query Match             39.5%;  Score 919;  DB 9;  Length 720;
  Best Local Similarity   98.4%;  Pred. No. 2.2e-62;
  Matches  180;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 588

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 648

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        649 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAEQK 708

Qy        207 ASS 209
                |
Db        709 LIS 711


RESULT 32
US-09-934-060A-4
; Sequence 4, Application US/09934060A
; Patent No. US20020155121A1
; GENERAL INFORMATION:
;  APPLICANT: DeVico, Anthony L.
;  APPLICANT:  Fouts, Timothy R.
;  APPLICANT:  Tuskan, Robert G.
;  TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE
;  FILE REFERENCE: 4115-144 CIP
;  CURRENT APPLICATION NUMBER: US/09/934,060A
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: US 09/684,026
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: US 60/158,321
;  PRIOR FILING DATE: 1999-10-08
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 720
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthesized construct
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (716)..(716)
;   OTHER INFORMATION: Xaa can be any amino acid
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (719)..(719)
;   OTHER INFORMATION: Xaa can be any amino acid
US-09-934-060A-4

  Query Match             39.5%;  Score 919;  DB 9;  Length 720;
  Best Local Similarity   98.4%;  Pred. No. 2.2e-62;
  Matches  180;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 588

Qy         87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 648

Qy        147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQK 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        649 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAEQK 708

Qy        207 ASS 209
                |
Db        709 LIS 711


RESULT 33
US-09-934-060A-26
; Sequence 26, Application US/09934060A
; Patent No. US20020155121A1
; GENERAL INFORMATION:
;  APPLICANT: DeVico, Anthony L.
;  APPLICANT:  Fouts, Timothy R.
;  APPLICANT:  Tuskan, Robert G.
;  TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE
;  FILE REFERENCE: 4115-144 CIP
;  CURRENT APPLICATION NUMBER: US/09/934,060A
;  CURRENT FILING DATE:  2001-08-21
;  PRIOR APPLICATION NUMBER: US 09/684,026
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: US 60/158,321
;  PRIOR FILING DATE: 1999-10-08
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
;   LENGTH: 178
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthesized construct
US-09-934-060A-26

  Query Match             39.2%;  Score 911;  DB 9;  Length 178;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-62;
```

```
  Matches  177;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          27 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 86
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           2 KVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSL 61

Qy          87 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 146
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          62 WDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTLTLESP 121

Qy         147 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 203
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         122 PGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLA 178


RESULT 34
US-10-024-329-33
; Sequence 33, Application US/10024329
; Publication No. US20030157063A1
; GENERAL INFORMATION:
;  APPLICANT: SANHADJI, Kamel
;  APPLICANT:  TOURAINE, Jean-Louis
;  APPLICANT:  LEROY, Pierre
;  APPLICANT:  MEHTALI, Majid
;  TITLE OF INVENTION: Gene therapy using anti-gp41 antibody and cd4 immunoadhesin
;  FILE REFERENCE: 109993
;  CURRENT APPLICATION NUMBER: US/10/024,329
;  CURRENT FILING DATE:  2001-12-21
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
;   LENGTH: 184
;   TYPE: PRT
;   ORGANISM: L ,D1, J1, D2 domains of human sCD4
US-10-024-329-33

  Query Match            37.9%;  Score 881;  DB 14;  Length 184;
  Best Local Similarity   95.3%;  Pred. No. 3e-60;
  Matches  181;  Conservative    1;  Mismatches    2;  Indels    6;  Gaps    5;

Qy           1 MNRGVPFRHLLLVLQLALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 60
               ||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||
Db           1 MNRGVPF-HLLLVLQLALLPAATQGKKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIK 59

Qy          61 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQL 120
               |||||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db          60 ILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC--VDQKEEVQL 117

Qy         121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSG 180
               |||||||||||||||||||||||||||||||||||| ||| |||||||| |||||||
Db         118 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPR-KNI-GGKTLSVS-LELQDSG 174

Qy         181 TWTCTVLQNQ 190
               ||||||||||:
Db         175 TWTCTVLQNK 184


RESULT 35
US-09-891-119A-10
; Sequence 10, Application US/09891119A
; Publication No. US20040013683A1
; GENERAL INFORMATION:
;  APPLICANT: Maddon, Paul J.
;  TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4
;  FILE REFERENCE: 24577-CY-B
;  CURRENT APPLICATION NUMBER: US/09/891,119A
;  CURRENT FILING DATE:  2001-06-25
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
;   LENGTH: 94
;   TYPE: PRT
;   ORGANISM: human
US-09-891-119A-10

  Query Match            21.2%;  Score 494;  DB 11;  Length 94;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-30;
  Matches   94;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          24 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 83
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSR 60

Qy          84 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEE 117
               |||||||||||||||||||||||||||||||||
Db          61 RSLWDQGNFPLIIKNLKIEDSDTYICEVEDQKEE 94


RESULT 36
US-10-105-545-26
; Sequence 26, Application US/10105545
; Publication No. US20030144479A1
; GENERAL INFORMATION:
;  APPLICANT: Mark, Greene I.
;  APPLICANT:  Williams, William V.
;  APPLICANT:  Weiner, David B.
;  APPLICANT:  Cohen, Jeffery  A.
;  APPLICANT:  Kieber-Emmons, Thomas
;  APPLICANT:  Williams, Robert M.
;  TITLE OF INVENTION: BIOLOGICALLY ACTIVE COMPOUNDS AND METHODS OF CONSTRUCTING AND USI
;  TITLE OF INVENTION:  SAME
;  FILE REFERENCE: 4040/1L492US2
;  CURRENT APPLICATION NUMBER: US/10/105,545
;  CURRENT FILING DATE:  2002-09-06
;  PRIOR APPLICATION NUMBER: US 08/752,816
;  PRIOR FILING DATE: 1996-11-21
;  PRIOR APPLICATION NUMBER: US 07/940,654
;  PRIOR FILING DATE: 1992-09-03
;  PRIOR APPLICATION NUMBER: US 07/702,833
;  PRIOR FILING DATE: 1991-05-20
;  PRIOR APPLICATION NUMBER: US 07/326,328
;  PRIOR FILING DATE: 1989-03-21
;  PRIOR APPLICATION NUMBER: US 07/074,264
;  PRIOR FILING DATE: 1987-07-16
;  PRIOR APPLICATION NUMBER: US 07/462,542
;  PRIOR FILING DATE: 1990-01-09
;  PRIOR APPLICATION NUMBER: US 07/648,303
;  PRIOR FILING DATE: 1991-01-25
;  PRIOR APPLICATION NUMBER: US 07/685,881
;  PRIOR FILING DATE: 1991-04-15
;  PRIOR APPLICATION NUMBER: US 07/574,391
;  PRIOR FILING DATE: 1990-08-27
;  PRIOR APPLICATION NUMBER: US 07/194,026
;  PRIOR FILING DATE: 1988-05-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
;   LENGTH: 93
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic peptide
US-10-105-545-26

  Query Match            20.9%;  Score 486;  DB 14;  Length 93;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-30;
  Matches   93;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          17 ALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKL 76
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ALLPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKL 60

Qy          77 NDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC 109
               |||||||||||||||||||||||||||||||||
Db          61 NDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC 93
```

```
RESULT 37
US-09-929-924-16
; Sequence 16, Application US/09929924
; Patent No. US20020147151A1
; GENERAL INFORMATION:
;  APPLICANT: Center, David M.
;  APPLICANT:  Cruickshank, William W.
;  APPLICANT:  Kornfeld, Hardy
;  TITLE OF INVENTION: IL-16 ANTAGONISTS
;  FILE REFERENCE: Research Corporation Tech., Inc.
;  CURRENT APPLICATION NUMBER: US/09/929,924
;  CURRENT FILING DATE:  2001-08-15
;  PRIOR APPLICATION NUMBER: 09/368,632
;  PRIOR FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
;   LENGTH: 73
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:IL-16 antagonist peptide
US-09-929-924-16

  Query Match             16.0%;  Score 373;  DB 9;  Length 73;
  Best Local Similarity   97.2%;  Pred. No. 2e-21;
  Matches   70;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        325 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNI 384
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db          1 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLLSDSGQVLLESNI 60

Qy        385 KVLPTWSTPVHP 396
              |||||||||| |
Db         61 KVLPTWSTPVQP 72


RESULT 38
US-09-929-924-36
; Sequence 36, Application US/09929924
; Patent No. US20020147151A1
; GENERAL INFORMATION:
;  APPLICANT: Center, David M.
;  APPLICANT:  Cruickshank, William W.
;  APPLICANT:  Kornfeld, Hardy
;  TITLE OF INVENTION: IL-16 ANTAGONISTS
;  FILE REFERENCE: Research Corporation Tech., Inc.
;  CURRENT APPLICATION NUMBER: US/09/929,924
;  CURRENT FILING DATE:  2001-08-15
;  PRIOR APPLICATION NUMBER: 09/368,632
;  PRIOR FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
;   LENGTH: 73
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:IL-16 antagonist peptide
US-09-929-924-36

  Query Match             16.0%;  Score 371;  DB 9;  Length 73;
  Best Local Similarity   95.8%;  Pred. No. 2.9e-21;
  Matches   69;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        325 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNI 384
              |||||||||||||||||||||||||||||||| |||||||||||||:|||||||||||||
Db          1 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCILSDSGQVLLESNI 60

Qy        385 KVLPTWSTPVHP 396
              |||||||||| |
Db         61 KVLPTWSTPVQP 72


RESULT 39
US-09-929-924-37
; Sequence 37, Application US/09929924
; Patent No. US20020147151A1
; GENERAL INFORMATION:
;  APPLICANT: Center, David M.
;  APPLICANT:  Cruickshank, William W.
;  APPLICANT:  Kornfeld, Hardy
;  TITLE OF INVENTION: IL-16 ANTAGONISTS
;  FILE REFERENCE: Research Corporation Tech., Inc.
;  CURRENT APPLICATION NUMBER: US/09/929,924
;  CURRENT FILING DATE:  2001-08-15
;  PRIOR APPLICATION NUMBER: 09/368,632
;  PRIOR FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
;   LENGTH: 73
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:IL-16 antagonist peptide
US-09-929-924-37

  Query Match             16.0%;  Score 371;  DB 9;  Length 73;
  Best Local Similarity   95.8%;  Pred. No. 2.9e-21;
  Matches   69;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        325 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNI 384
              |||||||||||||||||||||||||||||||| |||||||||||||:|||||||||||||
Db          1 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLISDSGQVLLESNI 60

Qy        385 KVLPTWSTPVHP 396
              |||||||||| |
Db         61 KVLPTWSTPVQP 72


RESULT 40
US-09-929-924-33
; Sequence 33, Application US/09929924
; Patent No. US20020147151A1
; GENERAL INFORMATION:
;  APPLICANT: Center, David M.
;  APPLICANT:  Cruickshank, William W.
;  APPLICANT:  Kornfeld, Hardy
;  TITLE OF INVENTION: IL-16 ANTAGONISTS
;  FILE REFERENCE: Research Corporation Tech., Inc.
;  CURRENT APPLICATION NUMBER: US/09/929,924
;  CURRENT FILING DATE:  2001-08-15
;  PRIOR APPLICATION NUMBER: 09/368,632
;  PRIOR FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
;   LENGTH: 73
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:IL-16 antagonist peptide
US-09-929-924-33

  Query Match             15.8%;  Score 368;  DB 9;  Length 73;
  Best Local Similarity   95.8%;  Pred. No. 4.9e-21;
  Matches   69;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        325 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNI 384
              |||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db          1 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWNCLLSDSGQVLLESNI 60

Qy        385 KVLPTWSTPVHP 396
```

```
              ||||||||||| |
Db         61 KVLPTWSTPVQP 72


RESULT 41
US-09-929-924-34
; Sequence 34, Application US/09929924
; Patent No. US20020147151A1
; GENERAL INFORMATION:
;  APPLICANT: Center, David M.
;  APPLICANT:  Cruickshank, William W.
;  APPLICANT:  Kornfeld, Hardy
;  TITLE OF INVENTION: IL-16 ANTAGONISTS
;  FILE REFERENCE: Research Corporation Tech., Inc.
;  CURRENT APPLICATION NUMBER: US/09/929,924
;  CURRENT FILING DATE:  2001-08-15
;  PRIOR APPLICATION NUMBER: 09/368,632
;  PRIOR FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
;   LENGTH: 73
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:IL-16 antagonist peptide
US-09-929-924-34

  Query Match             15.8%;  Score 367;  DB 9;  Length 73;
  Best Local Similarity   95.8%;  Pred. No. 5.9e-21;
  Matches   69;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        325 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNI 384
              ||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||
Db          1 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCSLSDSGQVLLESNI 60

Qy        385 KVLPTWSTPVHP 396
              ||||||||||| |
Db         61 KVLPTWSTPVQP 72


RESULT 42
US-09-929-924-35
; Sequence 35, Application US/09929924
; Patent No. US20020147151A1
; GENERAL INFORMATION:
;  APPLICANT: Center, David M.
;  APPLICANT:  Cruickshank, William W.
;  APPLICANT:  Kornfeld, Hardy
;  TITLE OF INVENTION: IL-16 ANTAGONISTS
;  FILE REFERENCE: Research Corporation Tech., Inc.
;  CURRENT APPLICATION NUMBER: US/09/929,924
;  CURRENT FILING DATE:  2001-08-15
;  PRIOR APPLICATION NUMBER: 09/368,632
;  PRIOR FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
;   LENGTH: 73
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:IL-16 antagonist peptide
US-09-929-924-35

  Query Match             15.8%;  Score 367;  DB 9;  Length 73;
  Best Local Similarity   95.8%;  Pred. No. 5.9e-21;
  Matches   69;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        325 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNI 384
              ||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||
Db          1 NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLSSDSGQVLLESNI 60

Qy        385 KVLPTWSTPVHP 396
              ||||||||||| |
Db         61 KVLPTWSTPVQP 72


RESULT 43
US-09-135-238B-8
; Sequence 8, Application US/09135238B
; Patent No. US20020177565A1
; GENERAL INFORMATION:
;  APPLICANT:  No. US20020177565A1an, Garry P.
;  APPLICANT:  Hitoshi, Yasumichi
;  TITLE OF INVENTION:  TOSO
;  FILE REFERENCE:  A65635-1/DJB/RMS
;  CURRENT APPLICATION NUMBER: US/09/135,238B
;  CURRENT FILING DATE:  1998-08-17
;  PRIOR APPLICATION NUMBER:  60/066,063
;  PRIOR FILING DATE:  1997-11-17
;  NUMBER OF SEQ ID NOS:  31
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 84
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (11)..(14)
;   OTHER INFORMATION: The xaa at positions 11 through 14 represents an
;   OTHER INFORMATION: unknown amino acid.
;   NAME/KEY: UNSURE
;   LOCATION: (28)
;   OTHER INFORMATION: The xaa at position 28 represents an unknown amino
;   OTHER INFORMATION: acid.
;   NAME/KEY: UNSURE
;   LOCATION: (41)
;   OTHER INFORMATION: The xaa at position 41 represents an unknown amino
;   OTHER INFORMATION: acid.
;   NAME/KEY: UNSURE
;   LOCATION: (46)..(48)
;   OTHER INFORMATION: The xaa at positions 46 through 48 represents an
;   OTHER INFORMATION: unknown amino acid.
;   NAME/KEY: UNSURE
;   LOCATION: (62)
;   OTHER INFORMATION: The xaa at position 62 represents an unknown amino
;   OTHER INFORMATION: acid.
US-09-135-238B-8

  Query Match             14.5%;  Score 337;  DB 9;  Length 84;
  Best Local Similarity   88.1%;  Pred. No. 1.5e-18;
  Matches   74;  Conservative    0;  Mismatches    0;  Indels   10;  Gaps    5;

Qy         37 VELTCTASQK----KSIQFHWKNSNQI-KILGNQGSFLTK-GPSK---LNDRADSRRSLW 87
              ||||||||||    |||||||||||||| ||||||||||| ||||   ||||||||||||
Db          1 VELTCTASQKXXXXKSIQFHWKNSNQIXKILGNQGSFLTKXGPSKXXXLNDRADSRRSLW 60

Qy         88 D-QGNFPLIIKNLKIEDSDTYICE 110
              | ||||||||||||||||||||||
Db         61 DXQGNFPLIIKNLKIEDSDTYICE 84


RESULT 44
US-09-997-165-2
; Sequence 2, Application US/09997165
; Patent No. US20020141999A1
; GENERAL INFORMATION:
;  APPLICANT: Lyman, Stewart D.
;  APPLICANT:  Fanslow, William C.
;  TITLE OF INVENTION: LIGAND FOR CD7 AND METHODS OF USE THEREOF
;  FILE REFERENCE: 2913-US
;  CURRENT APPLICATION NUMBER: US/09/997,165
;  CURRENT FILING DATE:  2001-11-27
;  PRIOR APPLICATION NUMBER: PCT/US00/14612
;  PRIOR FILING DATE: 2000-05-26
```

```
;  PRIOR APPLICATION NUMBER: 60/136,450
;  PRIOR FILING DATE: 1999-05-28
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 240
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-997-165-2

  Query Match             12.6%;  Score 293;  DB 9;  Length 240;
  Best Local Similarity   40.3%;  Pred. No. 1.6e-14;
  Matches   91;  Conservative   18;  Mismatches   49;  Indels   68;  Gaps    9;

Qy        270 PKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCE 329
              |:| :   ||| | | : ||          || |:  :       | : |:    |:||
Db          5 PRLLL---LPLLLALARGLPG---------ALAAQEVQQSPHCTTVPVGAS---VNITCS 49

Qy        330 VWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAG---------------------- 366
              |     :  |             :|  ::  | |
Db         50 TSGGLRGIYLRQL----------GPQPQDIIYYEDGVVPTTDRRFRGRIDFSGSQDNLT 98

Qy        367 --MWQCLLSDSG----QVLLESNI-----KVLPT------W--STPVHPRASALPAPPTG 407
                 | :  |||:|    | : | |:      || |      |   :   ||||||||||||
Db         99 ITMHRLQLSDTGTYTCQAITEVNVYGSGTLVLVTEEQSQGWHRCSDAPPRASALPAPPTG 158

Qy        408 SALPDPQTASALPDPPAASALPAALAVISFLLGLGLGVACVLARTR 453
              |||||||||||||||||||||||||||||||||||||||||||||:
Db        159 SALPDPQTASALPDPPAASALPAALAVISFLLGLGLGVACVLARTQ 204


RESULT 45
US-09-939-537-35
; Sequence 35, Application US/09939537
; Publication No. US20030138410A1
;   GENERAL INFORMATION:
;        APPLICANT: Seed, Brian
;                   Banapour, Babak
;                   Romeo, Charles
;                   Kolanus, Waldemar
;        TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
;                            CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
;        NUMBER OF SEQUENCES: 53
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Clark & Elbing LLP
;             STREET: 176 Federal Street
;             CITY: Boston
;             STATE: MA
;             COUNTRY: USA
;             ZIP: 02110
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette
;             COMPUTER: IBM Compatible
;             OPERATING SYSTEM: DOS
;             SOFTWARE: FastSEQ for Windows Version 2.0
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/939,537
;             FILING DATE: 24-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/284,391
;             FILING DATE: 02-AUG-1994
;             APPLICATION NUMBER: 08/195,395
;             FILING DATE: 14-FEB-1994
;             APPLICATION NUMBER: 07/847,566
;             FILING DATE: 06-MAR-1992
;             APPLICATION NUMBER: 07/665,961
;             FILING DATE: 07-MAR-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Elbing, Karen L
;             REGISTRATION NUMBER: 35,238
;             REFERENCE/DOCKET NUMBER: 00786/247001
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 617-428-0200
;             TELEFAX: 617-428-7045
;             TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 35:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 58 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-939-537-35

  Query Match             12.4%;  Score 288;  DB 10;  Length 58;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-15;
  Matches   58;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        396 PRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGVACVLARTR 453
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVISFLLGLGLGVACVLARTR 58


RESULT 46
US-10-105-545-25
; Sequence 25, Application US/10105545
; Publication No. US20030144479A1
; GENERAL INFORMATION:
;  APPLICANT: Mark, Greene I.
;  APPLICANT:  Williams, William V.
;  APPLICANT:  Weiner, David B.
;  APPLICANT:  Cohen, Jeffery  A.
;  APPLICANT:  Kieber-Emmons, Thomas
;  APPLICANT:  Williams, Robert M.
;  TITLE OF INVENTION: BIOLOGICALLY ACTIVE COMPOUNDS AND METHODS OF CONSTRUCTING AND USIN
;  TITLE OF INVENTION:  SAME
;  FILE REFERENCE: 4040/1L492US2
;  CURRENT APPLICATION NUMBER: US/10/105,545
;  CURRENT FILING DATE:  2002-09-06
;  PRIOR APPLICATION NUMBER: US 08/752,816
;  PRIOR FILING DATE: 1996-11-21
;  PRIOR APPLICATION NUMBER: US 07/940,654
;  PRIOR FILING DATE: 1992-09-03
;  PRIOR APPLICATION NUMBER: US 07/702,833
;  PRIOR FILING DATE: 1991-05-20
;  PRIOR APPLICATION NUMBER: US 07/326,328
;  PRIOR FILING DATE: 1989-03-21
;  PRIOR APPLICATION NUMBER: US 07/074,264
;  PRIOR FILING DATE: 1987-07-16
;  PRIOR APPLICATION NUMBER: US 07/462,542
;  PRIOR FILING DATE: 1990-01-09
;  PRIOR APPLICATION NUMBER: US 07/648,303
;  PRIOR FILING DATE: 1991-01-25
;  PRIOR APPLICATION NUMBER: US 07/685,881
;  PRIOR FILING DATE: 1991-04-15
;  PRIOR APPLICATION NUMBER: US 07/574,391
;  PRIOR FILING DATE: 1990-08-27
;  PRIOR APPLICATION NUMBER: US 07/194,026
;  PRIOR FILING DATE: 1988-05-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
;   LENGTH: 95
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: synthetic peptide
US-10-105-545-25

  Query Match             10.2%;  Score 238;  DB 14;  Length 95;
  Best Local Similarity   54.3%;  Pred. No. 8.4e-11;
  Matches   51;  Conservative   14;  Mismatches   25;  Indels    4;  Gaps    3;
```

```
Qy         19 LPAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQG-SFLTKG--PSK 75
Db          3 LLAVTQGKTLVLGKEGESAELPCESSQKKITVFTWKFSDQRKILGQHGKGVLIRGGSPSQ 62

Qy         76 LNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYIC 109
Db         63 F-DRFDSKKGAWEKGSFPLIINKLKMEDSQTYIC 95


RESULT 47
US-09-929-924-17
; Sequence 17, Application US/09929924
; Patent No. US20020147151A1
; GENERAL INFORMATION:
;  APPLICANT: Center, David M.
;  APPLICANT:  Cruickshank, William W.
;  APPLICANT:  Kornfeld, Hardy
;  TITLE OF INVENTION: IL-16 ANTAGONISTS
;  FILE REFERENCE: Research Corporation Tech., Inc.
;  CURRENT APPLICATION NUMBER: US/09/929,924
;  CURRENT FILING DATE:  2001-08-15
;  PRIOR APPLICATION NUMBER: 09/368,632
;  PRIOR FILING DATE: 1999-08-05
;  NUMBER OF SEQ ID NOS: 41
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
;   LENGTH: 72
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:IL-16 antagonist peptide
US-09-929-924-17

  Query Match             8.5%;  Score 197;  DB 9;  Length 72;
  Best Local Similarity   59.7%;  Pred. No. 8.7e-08;
  Matches   37;  Conservative   13;  Mismatches   12;  Indels    0;  Gaps    0;

Qy        326 LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLESNIK 385
Db          2 LTCEVMGPTSPKMRLTLKQENQEARVSEEQKVVQVVAPETGLWQCLLSEGDKVKMDSRIQ 61

Qy        386 VL 387
Db         62 VL 63


RESULT 48
US-10-311-823-13
; Sequence 13, Application US/10311823
; Publication No. US20040116683A1
; GENERAL INFORMATION:
;  APPLICANT: Bryan J. Boyle
;  APPLICANT:  Nancy Mize
;  APPLICANT:  Matthew Arterburn
;  APPLICANT:  Y. Tom Tang
;  APPLICANT:  George Yeung
;  APPLICANT:  Ping Zhou
;  APPLICANT:  Chenghua Liu
;  APPLICANT:  Vinod Asundi
;  APPLICANT:  Radoje T. Drmanac
;  APPLICANT:  Meng-Yun Wang
;  APPLICANT:  Lichuan Chen
;  APPLICANT:  Yea-Huey Yang
;  TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NEUROTRIMIN-LIKE
;  TITLE OF INVENTION:  POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: HYS-17CIP/US
;  CURRENT APPLICATION NUMBER: US/10/311,823
;  CURRENT FILING DATE:  2003-09-29
;  PRIOR APPLICATION NUMBER: PCT/US01/03651
;  PRIOR FILING DATE: 2001-02-02
;  PRIOR APPLICATION NUMBER: 09/632,085
;  PRIOR FILING DATE: 2000-08-02
;  PRIOR APPLICATION NUMBER: 09/560,875
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 570
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-311-823-13

  Query Match             7.0%;  Score 162;  DB 16;  Length 570;
  Best Local Similarity   22.1%;  Pred. No. 0.00077;
  Matches  108;  Conservative   77;  Mismatches  185;  Indels  118;  Gaps   25;

Qy          4 GVPFRHLLLVLQLALLPAATQGNKVVLGKKGD-----TVELTCTASQKKSIQFHWKNSNQ 58
Db         99 GVPIKSIRVDVQYLDEPMLTV-HQTVSDVRGNFYQEKTVFLRCTVNSNPPARFIWKRGSD 157

Qy         59 IKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI-IKNLKIEDSDTYICEVEDQKE- 116
Db        158 ----------TLSHSQDNG-VDIYEPLYTQGETKVLKLKNLRPQDYASYTCQVSVRNVC 205

Qy        117 --EVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQC-----------------R 156
Db        206 GIPDKAITFRLTNTTAPPALK-LSVNETLVVNPGENVTVQCLLTGGDPLPQLQWSHGPGP 264

Qy        157 SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN----QKKVEFKIDIVVLAFQKASSIVY 212
Db        265 LPLGALAQGG-TLSIPSVQARDSGYYNCTATNNVGNPAKKT---VNLLVRSMKNATFQIT 320

Qy        213 K---KEGEQVEFSFPL-------AFTVEKLTGSGELWWQ-AERASSSKSWITFDLKNKEV 261
Db        321 PDVIKESENIQLGQDLKLSCHVDAVPQEKVTYQ---WFKNGKPARMSKRLL---VTRNDP 374

Qy        262 SVKRVTQDPKLQMGKKLPLHL----------TLPQA-LPQYAGSGNLTLALEAKTGKLHQ 310
Db        375 ELPAVTSSLEL-----IDLHFSDYGTYLCMASFPGAPVPDLSVEVNISSETVPPTISVPK 429

Qy        311 EVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAK-----VSKREKPVWVLNPE- 364
Db        430 GRAVVTVREGS-PAELQCEVRGKPRPPVLWS--RVDKEAALLPSGLPLEETPDGKLRLER 486

Qy        365 -----AGMWQCLLSDSG-------QVLLESNIKV---------------LPTWSTPVHPR 397
Db        487 VSRDMSGTYRCQTARYNGFNVRPREAQVQLNVQCESTTPPGLPSQLSFWLPAGLSPLHPL 546

Qy        398 ASALPAPP 405
Db        547 THPSPTKP 554


RESULT 49
US-10-311-823-7
; Sequence 7, Application US/10311823
; Publication No. US20040116683A1
; GENERAL INFORMATION:
;  APPLICANT: Bryan J. Boyle
;  APPLICANT:  Nancy Mize
;  APPLICANT:  Matthew Arterburn
;  APPLICANT:  Y. Tom Tang
;  APPLICANT:  George Yeung
;  APPLICANT:  Ping Zhou
;  APPLICANT:  Chenghua Liu
;  APPLICANT:  Vinod Asundi
;  APPLICANT:  Radoje T. Drmanac
;  APPLICANT:  Meng-Yun Wang
;  APPLICANT:  Lichuan Chen
;  APPLICANT:  Yea-Huey Yang
;  TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NEUROTRIMIN-LIKE
;  TITLE OF INVENTION:  POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: HYS-17CIP/US
;  CURRENT APPLICATION NUMBER: US/10/311,823
```

```
;  CURRENT FILING DATE:  2003-09-29
;  PRIOR APPLICATION NUMBER: PCT/US01/03651
;  PRIOR FILING DATE: 2001-02-02
;  PRIOR APPLICATION NUMBER: 09/632,085
;  PRIOR FILING DATE: 2000-08-02
;  PRIOR APPLICATION NUMBER: 09/560,875
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 586
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-311-823-7

  Query Match             7.0%;  Score 162;  DB 16;  Length 586;
  Best Local Similarity   22.1%;  Pred. No. 0.0008;
  Matches  108;  Conservative   77;  Mismatches  185;  Indels  118;  Gaps   25;

Qy          4 GVPFRHLLLVLQLALLPAATQGNKVVLGKKGD-----TVELTCTASQKKSIQFHWKNSNQ 58
              ||| : : : :|    |  |  :: |   :|:     || | || :     :| ||  :
Db        115 GVPIKSIRVDVQYLDEPMLTV-HQTVSDVRGNFYQEKTVFLRCTVNSNPPARFIWKRGSD 173

Qy         59 IKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI-IKNLKIEDSDTYICEVEDQKE- 116
                  |   |: |   |    |: ||   :: :|||: :|  :| |:|  :
Db        174 ----------TLSHSQDNG-VDIYEPLYTQGETKVLKLKNLRPQDYASYTCQVSVRNVC 221

Qy        117 --EVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQC-----------------R 156
                : : | ||  :    |:  |:  ||   || : :|||
Db        222 GIPDKAITFRLTNTTAPPALK-LSVNETLVVNPGENVTVQCLLTGGDPLPQLQWSHGPGP 280

Qy        157 SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN----QKKVEFKIDIVVLAFQKASSIVY 212
              | |   ||| |||:  :: :||| : ||   |     ||    ::::| : : |:  :
Db        281 LPLGALAQGG-TLSIPSVQARDSGYYNCTATNNVGNPAKKT---VNLLVRSMKNATFQIT 336

Qy        213 K---KEGEQVEFSFPL-------AFTVEKLTGSGELWWQ-AERASSSKSWITFDLKNKEV 261
                || | ::    |       |   ||:|     |::  : |  || :   :   :
Db        337 PDVIKESENIQLGQDLKLSCHVDAVPQEKVTYQ---WFKNGKPARMSKRLL---VTRNDP 390

Qy        262 SVKRVTQDPKLQMGKKLPLHL----------TLPQA-LPQYAGSGNLTLALEAKTGKLHQ 310
              :  ||   :|     : ||          : | | :|  :   |::      |  : :
Db        391 ELPAVTSSLEL-----IDLHFSDYGTYLCMASFPGAPVPDLSVEVNISSETVPPTISVPK 445

Qy        311 EVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAK-----VSKREKPVWVLNPE- 364
              :| :|       | ||| |   | :: |    :|||      :   | |   | |
Db        446 GRAVVTVREGS-PAELQCEVRGKPRPPVLWS--RVDKEAALLPSGLPLEETPDGKLRLER 502

Qy        365 -----AGMWQCLLSDSG-------QVLLESNIKV--------------LPTWSTPVHPR 397
                  :| ::|  :          :  :: |::                ||   :|:||
Db        503 VSRDMSGTYRCQTARYNGFNVRPREAQVQLNVQCESTTPPGLPSQLSFWLPAGLSPLHPL 562

Qy        398 ASALPAPP 405
                | |
Db        563 THPSPTKP 570


RESULT 50
US-10-094-886-52
; Sequence 52, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
;  APPLICANT: Kekuda, Ramesh
;  APPLICANT:  Tchernev, Velizar T.
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Burgess, Catherine
;  APPLICANT:  Vernet, Corine A.
;  APPLICANT:  Li, Li
;  APPLICANT:  Gorman, Linda
;  APPLICANT:  Malyankar, Uriel M.
;  APPLICANT:  Boldog, Ferenc
;  APPLICANT:  Guo, Xiaojia
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Taupier, Raymond J., Jr.
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Pena, Carol
;  APPLICANT:  Gangolli, Esha
;  APPLICANT:  Gusev, Vladimir
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Gerlach, Valerie
;  APPLICANT:  Pochart, Pascal
;  APPLICANT:  Fernandes, Elma
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Spaderna, Steven
;  APPLICANT:  LaRochelle, William
;  APPLICANT:  Zhong, Mei
;  TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
;  FILE REFERENCE: 21402-290 B
;  CURRENT APPLICATION NUMBER: US/10/094,886
;  CURRENT FILING DATE:  2002-03-07
;  PRIOR APPLICATION NUMBER: 60/274,322
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/313,182
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 60/288,052
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: 60/318,510
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/274,281
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/314,018
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: 60/274,194
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/274,849
;  PRIOR FILING DATE: 2001-03-09
;  PRIOR APPLICATION NUMBER: 60/296,693
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: 60/313,626
;  PRIOR FILING DATE: 2001-08-21
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 298
;  SOFTWARE: PatentIn 2.1
; SEQ ID NO 52
;   LENGTH: 1315
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-094-886-52

  Query Match             6.8%;  Score 157;  DB 15;  Length 1315;
  Best Local Similarity   23.3%;  Pred. No. 0.0059;
  Matches  112;  Conservative   64;  Mismatches  181;  Indels  124;  Gaps   25;

Qy         36 TVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI 95
              || | || :     :| ||  :             |   |: |   |    |: ||   ::
Db        152 TVFLRCTVNSNPPARFIWKRGSD-----------TLSHSQDNG-VDIYEPLYTQGETKVL 199

Qy         96 -IKNLKIEDSDTYICEVEDQKE---EVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSP 151
                 :|||: :|  :| |:|  :       : : | ||  :    |:  |:  ||   || :
Db        200 KLKNLRPQDYASYTCQVSVRNVCGIPDKAITFRLTNTTAPPALK-LSVNETLVVNPGENV 258

Qy        152 SVQC-----------------RSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN---- 189
              :|||                | |   ||| |||:  :: :||| : ||   |
Db        259 TVQCLLTGGDPLPQLQWSHGPGPLPLGALAQGG-TLSIPSVQARDSGYYNCTATNNVGNP 317

Qy        190 QKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFSFPL-------AFTVEKLTGSGELW 239
               ||    ::::| : : |:  :     || | ::    |       |   ||:|     |
Db        318 AKKT---VNLLVRSMKNATFQITPDVIKESENIQLGQDLKLSCHVDAVPQEKVTYQ---W 371
```

```
Qy        240 WQ-AERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHL----------TLPQA- 287
              ::  : |  ||  :   :  :  :  ||   :|     : ||           : | |
Db        372 FKNGKPARMSKRLL---VTRNDPELPAVTSSLEL-----IDLHFSDYGTYLCMASFPGAP 423

Qy        288 LPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENK 347
              :|  :   |::     |  : :  :| :|      | ||| |   | :: |    :|
Db        424 VPDLSVEVNISSETVPPTISVPKGRAVVTVREGS-PAELQCEVRGKPRPPVLWS--RVDK 480

Qy        348 EAK-----VSKREKPVWVLNPE------AGMWQC---------LLSDSGQVLLE------ 381
              ||      :   | |   |  |      :| ::|         :     || |
Db        481 EAALLPSGLPLEETPDGKLRLERVSRDMSGTYRCQTARYNGFNVRPREAQVQLNVQFPPE 540

Qy        382 ---SNIKVLPTWSTPVHPRASALPAPP----------TGSALPDPQTASALPDPPAASAL 428
                 |:  |      ||  | | |  |           | || |       |  |||:
Db        541 VEPSSQDVRQALGRPVLLRCSLLRGSPQRIASAVWRFKGQLLPPP------PVVPAAAEA 594

Qy        429 P 429
              |
Db        595 P 595
```

```
RESULT 51
US-10-094-886-38
; Sequence 38, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
;  APPLICANT: Kekuda, Ramesh
;  APPLICANT:  Tchernev, Velizar T.
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Spytek, Kimberly A.
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Burgess, Catherine
;  APPLICANT:  Vernet, Corine A.
;  APPLICANT:  Li, Li
;  APPLICANT:  Gorman, Linda
;  APPLICANT:  Malyankar, Uriel M.
;  APPLICANT:  Boldog, Ferenc
;  APPLICANT:  Guo, Xiaojia
;  APPLICANT:  Shenoy, Suresh
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Taupier, Raymond J., Jr.
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Casman, Stacie
;  APPLICANT:  Pena, Carol
;  APPLICANT:  Gangolli, Esha
;  APPLICANT:  Gusev, Vladimir
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Gerlach, Valerie
;  APPLICANT:  Pochart, Pascal
;  APPLICANT:  Fernandes, Elma
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Spaderna, Steven
;  APPLICANT:  LaRochelle, William
;  APPLICANT:  Zhong, Mei
;  TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
;  FILE REFERENCE: 21402-290 B
;  CURRENT APPLICATION NUMBER: US/10/094,886
;  CURRENT FILING DATE:  2002-03-07
;  PRIOR APPLICATION NUMBER: 60/274,322
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/313,182
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 60/288,052
;  PRIOR FILING DATE: 2001-05-02
;  PRIOR APPLICATION NUMBER: 60/318,510
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/274,281
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/314,018
;  PRIOR FILING DATE: 2001-08-21
;  PRIOR APPLICATION NUMBER: 60/274,194
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/274,849
;  PRIOR FILING DATE: 2001-03-09
;  PRIOR APPLICATION NUMBER: 60/296,693
;  PRIOR FILING DATE: 2001-06-07
;  PRIOR APPLICATION NUMBER: 60/313,626
;  PRIOR FILING DATE: 2001-08-21
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 298
;  SOFTWARE: PatentIn 2.1
; SEQ ID NO 38
;   LENGTH: 1386
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-094-886-38

  Query Match             6.8%;  Score 157;  DB 15;  Length 1386;
  Best Local Similarity   23.3%;  Pred. No. 0.0063;
  Matches  112;  Conservative   64;  Mismatches  181;  Indels  124;  Gaps   25;

Qy         36 TVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI 95
              || | || :     :| ||  :             |   |: |   |     |: ||   ::
Db        152 TVFLRCTVNSNPPARFIWKRGSD-----------TLSHSQDNG-VDIYEPLYTQGETKVL 199

Qy         96 -IKNLKIEDSDTYICEVEDQKE---EVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSP 151
               :|||: :|  :| |:|  :       : : | ||  :    |:  |:  ||   || :
Db        200 KLKNLRPQDYASYTCQVSVRNVCGIPDKAITFRLTNTTAPPALK-LSVNETLVVNPGENV 258

Qy        152 SVQC-----------------RSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN---- 189
              :|||                | |   ||| |||:  :: :||| : ||   |
Db        259 TVQCLLTGGDPLPQLQWSHGPGPLPLGALAQGG-TLSIPSVQARDSGYYNCTATNNVGNP 317

Qy        190 QKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFSFPL-------AFTVEKLTGSGELW 239
               ||    ::::| : : |:  :     || | ::    |       |   ||:|     |
Db        318 AKKT---VNLLVRSMKNATFQITPDVIKESENIQLGQDLKLSCHVDAVPQEKVTYQ---W 371

Qy        240 WQ-AERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHL----------TLPQA- 287
              ::  : |  ||  :   :   :  :  ||   :|     : ||           : | |
Db        372 FKNGKPARMSKRLL---VTRNDPELPAVTSSLEL-----IDLHFSDYGTYLCMASFPGAP 423

Qy        288 LPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENK 347
              :|  :   |::     |  : :  :| :|      | ||| |   | :: |    :|
Db        424 VPDLSVEVNISSETVPPTISVPKGRAVVTVREGS-PAELQCEVRGKPRPPVLWS--RVDK 480

Qy        348 EAK-----VSKREKPVWVLNPE------AGMWQC---------LLSDSGQVLLE------ 381
              ||      :   | |   |  |      :| ::|         :     || |
Db        481 EAALLPSGLPLEETPDGKLRLERVSRDMSGTYRCQTARYNGFNVRPREAQVQLNVQFPPE 540

Qy        382 ---SNIKVLPTWSTPVHPRASALPAPP----------TGSALPDPQTASALPDPPAASAL 428
                 |:  |      ||  | | |  |           | || |       |  |||:
Db        541 VEPSSQDVRQALGRPVLLRCSLLRGSPQRIASAVWRFKGQLLPPP------PVVPAAAEA 594

Qy        429 P 429
              |
Db        595 P 595
```

```
RESULT 52
US-10-047-542-69
; Sequence 69, Application US/10047542
; Publication No. US20020168367A1
; GENERAL INFORMATION:
;  APPLICANT: LARRICK, JAMES W.
;  APPLICANT:  WYCOFF, KEITH L.
;  TITLE OF INVENTION: NOVEL IMMUNOADHESINS FOR TREATING AND PREVENTING VIRAL
;  TITLE OF INVENTION:  AND BACTERIAL DISEASES
;  FILE REFERENCE: 030905.0004.CIP1
;  CURRENT APPLICATION NUMBER: US/10/047,542
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: PCT/US01/13932
;  PRIOR FILING DATE: 2001-04-28
;  PRIOR APPLICATION NUMBER: 60/200,298
;  PRIOR FILING DATE: 2000-04-28
```

```
;  NUMBER OF SEQ ID NOS: 101
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-047-542-69

  Query Match             6.7%;  Score 155;  DB 13;  Length 739;
  Best Local Similarity   21.9%;  Pred. No. 0.0038;
  Matches   91;  Conservative   71;  Mismatches  166;  Indels   88;  Gaps   18;

Qy          8 RHLLLVLQLALLPAAT--QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQ 65
              |  :  ||: : |  |    |     ::| :| :||::    : :  |
Db        211 RQAVKELQVYISPKNTVISVNPSTKLQEGGSVTMTCSSEGLPAPEIFW------------ 258

Qy         66 GSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICE----VEDQKEEVQLL 121
                  ||  |  : :      ||  | :  :::|||  |:||    :   ::||:|:
Db        259 --------SKKLDNGNLQHL---SGNATLTLIAMRMEDSGIYVCEGVNLIGKNRKEVELI 307

Qy        122 V----FGLTANSDTHLLQ--GQSLTLTLESPPGSSPSVQCRSPRGKNIQG-------GKT 168
              |    | :  :    :   | |: ||      |||   |:     : |        |
Db        308 VQEKPFTVEISPGPRIAAQIGDSVMLTCSVMGCESPSFSWRTQIDSPLSGKVRSEGTNST 367

Qy        169 LSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFS--- 222
              |::| :  ::  :: |||    ||:|  | : : :| :   |         |  |  |
Db        368 LTLSPVSFENEHSYLCTVTCGHKKLEKGIQVELYSFPRDPEIEMSGGLVNGSSVTVSCKV 427

Qy        223 ---FPL-AFTVEKLTGSGELWWQAERASSSKSWITF----DLKNKEVSVKRVTQDPKLQ- 273
                 :||    :| | |   |              | |    |:|: |    :|  | ::
Db        428 PSVYPLDRLEIELLKGETIL-----------ENIEFLEDTDMKSLENKSLEMTFIPTIED 476

Qy        274 MGKKL----PLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQK----N 325
               || |     ||:   :  |:   |  | |        ::  ::|  :: |::    |
Db        477 TGKALVCQAKLHIDDMEFEPKQRQS---TQTLYVNVAP--RDTTVLVSPSSILEEGSSVN 531

Qy        326 LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLE 381
              :||   |  :||:: | :| | | :       : ::: :       : |||  | |
Db        532 MTCLSQGFPAPKILWSRQLPNGELQPLSENATLTLISTK-------MEDSGVYLCE 580


RESULT 53
US-10-207-655-139
; Sequence 139, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
;  APPLICANT: Ledbetter, Jeffrey A.
;  APPLICANT:  Hayden-Ledbetter, Martha S.
;  TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
;  FILE REFERENCE: 390069.401C1
;  CURRENT APPLICATION NUMBER: US/10/207,655
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 426
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 139
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-207-655-139

  Query Match             6.7%;  Score 155;  DB 14;  Length 739;
  Best Local Similarity   21.9%;  Pred. No. 0.0038;
  Matches   91;  Conservative   71;  Mismatches  166;  Indels   88;  Gaps   18;

Qy          8 RHLLLVLQLALLPAAT--QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQ 65
              |  :  ||: : |  |    |     ::| :| :||::    : :  |
Db        211 RQAVKELQVYISPKNTVISVNPSTKLQEGGSVTMTCSSEGLPAPEIFW------------ 258

Qy         66 GSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICE----VEDQKEEVQLL 121
                  ||  |  : :      ||  | :  :::|||  |:||    :   ::||:|:
Db        259 --------SKKLDNGNLQHL---SGNATLTLIAMRMEDSGIYVCEGVNLIGKNRKEVELI 307

Qy        122 V----FGLTANSDTHLLQ--GQSLTLTLESPPGSSPSVQCRSPRGKNIQG-------GKT 168
              |    | :  :    :   | |: ||      |||   |:     : |        |
Db        308 VQEKPFTVEISPGPRIAAQIGDSVMLTCSVMGCESPSFSWRTQIDSPLSGKVRSEGTNST 367

Qy        169 LSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFS--- 222
              |::| :  ::  :: |||    ||:|  | : : :| :   |         |  |  |
Db        368 LTLSPVSFENEHSYLCTVTCGHKKLEKGIQVELYSFPRDPEIEMSGGLVNGSSVTVSCKV 427

Qy        223 ---FPL-AFTVEKLTGSGELWWQAERASSSKSWITF----DLKNKEVSVKRVTQDPKLQ- 273
                 :||    :| | |   |              | |    |:|: |    :|  | ::
Db        428 PSVYPLDRLEIELLKGETIL-----------ENIEFLEDTDMKSLENKSLEMTFIPTIED 476

Qy        274 MGKKL----PLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQK----N 325
               || |     ||:   :  |:   |  | |        ::  ::|  :: |::    |
Db        477 TGKALVCQAKLHIDDMEFEPKQRQS---TQTLYVNVAP--RDTTVLVSPSSILEEGSSVN 531

Qy        326 LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLE 381
              :||   |  :||:: | :| | | :       : ::: :       : |||  | |
Db        532 MTCLSQGFPAPKILWSRQLPNGELQPLSENATLTLISTK-------MEDSGVYLCE 580


RESULT 54
US-10-234-041-7
; Sequence 7, Application US/10234041
; Publication No. US20030153731A1
; GENERAL INFORMATION:
;  APPLICANT: Biogen, Inc.
;  APPLICANT:  Hession, Catherine A.
;  APPLICANT:  Lobb, Roy R.
;  APPLICANT:  Goelz, Susan E.
;  APPLICANT:  Osborn, Laurelee
;  APPLICANT:  Benjamin, Christopher D.
;  APPLICANT:  Rosa, Margaret D.
;  TITLE OF INVENTION: Endothelial Cell-Leukocyte Adhesion
;  TITLE OF INVENTION:  Molecules (ELAMs) and Molecules Involved in Leukocyte
;  TITLE OF INVENTION:  Adhesion (MILAs)
;  FILE REFERENCE: B124CP2DV2CN
;  CURRENT APPLICATION NUMBER: US/10/234,041
;  CURRENT FILING DATE:  2003-04-08
;  PRIOR APPLICATION NUMBER: 08/473,764
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: 08/342,642
;  PRIOR FILING DATE: 1994-11-21
;  PRIOR APPLICATION NUMBER: 07/608,298
;  PRIOR FILING DATE: 1990-10-31
;  PRIOR APPLICATION NUMBER: 07/452,675
;  PRIOR FILING DATE: 1989-12-18
;  PRIOR APPLICATION NUMBER: 07/359,516
;  PRIOR FILING DATE: 1989-06-01
;  PRIOR APPLICATION NUMBER: 07/345,151
;  PRIOR FILING DATE: 1989-04-28
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 739
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-234-041-7

  Query Match             6.7%;  Score 155;  DB 14;  Length 739;
  Best Local Similarity   21.9%;  Pred. No. 0.0038;
  Matches   91;  Conservative   71;  Mismatches  166;  Indels   88;  Gaps   18;

Qy          8 RHLLLVLQLALLPAAT--QGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQ 65
              |  :  ||: : |  |    |     ::| :| :||::    : :  |
Db        211 RQAVKELQVYISPKNTVISVNPSTKLQEGGSVTMTCSSEGLPAPEIFW------------ 258

Qy         66 GSFLTKGPSKLNDRADSRRSLWDQGNFPLIIKNLKIEDSDTYICE----VEDQKEEVQLL 121
                  ||  |  : :      ||  | :  :::|||  |:||    :   ::||:|:
Db        259 --------SKKLDNGNLQHL---SGNATLTLIAMRMEDSGIYVCEGVNLIGKNRKEVELI 307

Qy        122 V----FGLTANSDTHLLQ--GQSLTLTLESPPGSSPSVQCRSPRGKNIQG-------GKT 168
              |    | :  :    :   | |: ||      |||   |:     : |        |
Db        308 VQEKPFTVEISPGPRIAAQIGDSVMLTCSVMGCESPSFSWRTQIDSPLSGKVRSEGTNST 367

Qy        169 LSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFS--- 222
              |::| :  ::  :: |||    ||:|  | : : :| :   |         |  |  |
Db        368 LTLSPVSFENEHSYLCTVTCGHKKLEKGIQVELYSFPRDPEIEMSGGLVNGSSVTVSCKV 427

Qy        223 ---FPL-AFTVEKLTGSGELWWQAERASSSKSWITF----DLKNKEVSVKRVTQDPKLQ- 273
                 :||    :| | |   |              | |    |:|: |    :|  | ::
Db        428 PSVYPLDRLEIELLKGETIL-----------ENIEFLEDTDMKSLENKSLEMTFIPTIED 476

Qy        274 MGKKL----PLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQK----N 325
               || |     ||:   :  |:   |  | |        ::  ::|  :: |::    |
Db        477 TGKALVCQAKLHIDDMEFEPKQRQS---TQTLYVNVAP--RDTTVLVSPSSILEEGSSVN 531

Qy        326 LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLE 381
              :||   |  :||:: | :| | | :       : ::: :       : |||  | |
Db        532 MTCLSQGFPAPKILWSRQLPNGELQPLSENATLTLISTK-------MEDSGVYLCE 580
```

```
Qy        169 LSVSQLELQDSGTWTCTVLQNQKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFS--- 222
Db        368 LTLSPVSFENEHSYLCTVTCGHKKLEKGIQVELYSFPRDPEIEMSGGLVNGSSVTVSCKV 427

Qy        223 ---FPL-AFTVEKLTGSGELWWQAERASSSKSWITF----DLKNKEVSVKRVTQDPKLQ- 273
Db        428 PSVYPLDRLEIELLKGETIL-----------ENIEFLEDTDMKSLENKSLEMTFIPTIED 476

Qy        274 MGKKL----PLHLTLPQALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQK----N 325
Db        477 TGKALVCQAKLHIDDMEFEPKQRQS---TQTLYVNVAP--RDTTVLVSPSSILEEGSSVN 531

Qy        326 LTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQVLLE 381
Db        532 MTCLSQGFPAPKILWSRQLPNGELQPLSENATLTLISTK-------MEDSGVYLCE 580


RESULT 55
US-10-231-956A-325
; Sequence 325, Application US/10231956A
; Publication No. US20040053233A1
; GENERAL INFORMATION:
;  APPLICANT: Lorens, James B.
;  APPLICANT:  Xu, Weiduan
;  APPLICANT:  Bogenberger, Jakob
;  APPLICANT:  Holland, Sacha
;  APPLICANT:  Rigel Pharmaceuticals, Incorporated
;  TITLE OF INVENTION: Modulators of Angiogenesis
;  FILE REFERENCE: 021044-004100US
;  CURRENT APPLICATION NUMBER: US/10/231,956A
;  CURRENT FILING DATE:  2001-08-30
;  NUMBER OF SEQ ID NOS: 522
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 325
;   LENGTH: 1479
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-231-956A-325

  Query Match             6.5%;  Score 152;  DB 12;  Length 1479;
  Best Local Similarity   21.5%;  Pred. No. 0.017;
  Matches  100;  Conservative   57;  Mismatches  155;  Indels  154;  Gaps   19;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
Db        260 GNTVYFTCRAEGNPKPEIIWLRNN---------------NELSMKTDSRLNLLDDGT-- 301

Qy         94 LIIKNLKIEDSDTYICEV-----EDQKEEVQLLVFGLTAN-------SDTHLLQGQSLTL 141
Db        302 LMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTL 361

Qy        142 ----TLESPPGSSPSVQCRSP----RGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
Db        362 ECSATGHPPPRISWTRGDRTPLPVDPRVNITPSGGLYIQNVVQGDSGEYACSATNNIDSV 421

Qy        194 EFKIDIVVLAFQKASSIVYKK---EGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKS 250
Db        422 HATAFIIVQALPQFTVTPQDRVVIEGQTVDF------------------QCEAKGNPPP 462

Qy        251 WITFDLKNKEVSVKR---VTQDPKLQM--------------------GKKLPLHLTL-PQ 286
Db        463 VIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPR 522

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLEN 346
Db        523 VTPVFASIPSDTTV----------EVGANV--------QLPCSSQGEPEPAITW-----N 559

Qy        347 KEAKVSKREKPVWVLNPE------------AGMWQCLL-----SDSGQVLLESNIKVLPT 389
Db        560 KDG-VQVTESGKFHISPEGFLTINDVGPADAGRYECVARNTIGSASVSMVLSVNV----- 613

Qy        390 WSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVI 435
```

```
                                  ||    |   ||  |:::  |:| :
Db        614 --------------------PD---VSRNGDPFVATSIVEAIATV 635


RESULT 56
US-10-211-462-87
; Sequence 87, Application US/10211462
; Publication No. US20040033495A1
; GENERAL INFORMATION:
;  APPLICANT: Murray, Richard
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Angiogenesis Modulators
;  FILE REFERENCE: 018501-006200US
;  CURRENT APPLICATION NUMBER: US/10/211,462
;  CURRENT FILING DATE:  2003-02-13
;  PRIOR APPLICATION NUMBER: US 09/784,356
;  PRIOR FILING DATE: 2001-02-14
;  PRIOR APPLICATION NUMBER: US 09/791,390
;  PRIOR FILING DATE: 2001-02-22
;  PRIOR APPLICATION NUMBER: US 60/310,025
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/334,244
;  PRIOR FILING DATE: 2001-11-29
;  NUMBER OF SEQ ID NOS: 230
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 87
;   LENGTH: 1496
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-211-462-87

  Query Match             6.5%;  Score 152;  DB 12;  Length 1496;
  Best Local Similarity   21.5%;  Pred. No. 0.017;
  Matches  100;  Conservative   57;  Mismatches  155;  Indels  154;  Gaps   19;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
Db        277 GNTVYFTCRAEGNPKPEIIWLRNN---------------NELSMKTDSRLNLLDDGT-- 318

Qy         94 LIIKNLKIEDSDTYICEV-----EDQKEEVQLLVFGLTAN-------SDTHLLQGQSLTL 141
Db        319 LMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTL 378

Qy        142 ----TLESPPGSSPSVQCRSP----RGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
Db        379 ECSATGHPPPRISWTRGDRTPLPVDPRVNITPSGGLYIQNVVQGDSGEYACSATNNIDSV 438

Qy        194 EFKIDIVVLAFQKASSIVYKK---EGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKS 250
Db        439 HATAFIIVQALPQFTVTPQDRVVIEGQTVDF------------------QCEAKGNPPP 479

Qy        251 WITFDLKNKEVSVKR---VTQDPKLQM--------------------GKKLPLHLTL-PQ 286
Db        480 VIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPR 539

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLEN 346
Db        540 VTPVFASIPSDTTV----------EVGANV--------QLPCSSQGEPEPAITW-----N 576

Qy        347 KEAKVSKREKPVWVLNPE------------AGMWQCLL-----SDSGQVLLESNIKVLPT 389
Db        577 KDG-VQVTESGKFHISPEGFLTINDVGPADAGRYECVARNTIGSASVSMVLSVNV----- 630

Qy        390 WSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVI 435
                                  ||    |   ||  |:::  |:| :
Db        631 --------------------PD---VSRNGDPFVATSIVEAIATV 652


RESULT 57
```

```
US-10-021-660-125
; Sequence 125, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
;  APPLICANT: Murray, Richard
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  EOS Biotechnology, Inc.
;  TITLE OF INVENTION: No. US20030152926A1el Methods of Diagnosis of Angiogenesis,
;  TITLE OF INVENTION:  Compositions and Methods of Screening for Angiogenesis
;  TITLE OF INVENTION:  Modulators
;  FILE REFERENCE: 018501-000710US
;  CURRENT APPLICATION NUMBER: US/10/021,660
;  CURRENT FILING DATE:  2001-12-06
;  PRIOR APPLICATION NUMBER: US/09/784,356
;  PRIOR FILING DATE: 2001-02-14
;  PRIOR APPLICATION NUMBER: US 09/637,977
;  PRIOR FILING DATE: 2000-08-11
;  NUMBER OF SEQ ID NOS: 135
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 125
;   LENGTH: 1496
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-021-660-125

  Query Match             6.5%;  Score 152;  DB 14;  Length 1496;
  Best Local Similarity   21.5%;  Pred. No. 0.017;
  Matches  100;  Conservative   57;  Mismatches  155;  Indels  154;  Gaps   19;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |:||  || |      :  |  :|               ::|: : ||| :| | |
Db        277 GNTVYFTCRAEGNPKPEIIWLRNN---------------NELSMKTDSRLNLLDDGT-- 318

Qy         94 LIIKNLKIEDSDTYICEV-----EDQKEEVQLLVFGLTAN-------SDTHLLQGQSLTL 141
              |:|:| :  |   | |       | : :|| |  ||  |         :| :| |:|:||
Db        319 LMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTL 378

Qy        142 ----TLESPPGSSPSVQCRSP----RGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
                  |   ||  | :  |:|       ||     | :  :  ||| : |:   |   |
Db        379 ECSATGHPPPRISWTRGDRTPLPVDPRVNITPSGGLYIQNVVQGDSGEYACSATNNIDSV 438

Qy        194 EFKIDIVVLAFQKASSIVYKK---EGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKS 250
                |:| |  : :    :   ||: |:|                   | |   :
Db        439 HATAFIIVQALPQFTVTPQDRVVIEGQTVDF-----------------QCEAKGNPPP 479

Qy        251 WITFDLKNKEVSVKR---VTQDPKLQM-------------------GKKLPLHLTL-PQ 286
               | :    ::|| |   |      |::                        :|:  |||: |:
Db        480 VIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPR 539

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLEN 346
               | :|  : |          ||  |       | |   |  | :        |
Db        540 VTPVFASIPSDTTV----------EVGANV--------QLPCSSQGEPEPAITW-----N 576

Qy        347 KEAKVSKREKPVWVLNPE------------AGMWQCLL-----SDSGQVLLESNIKVLPT 389
              |:  |   |   : ::||              || ::|:      | |  ::|  |:
Db        577 KDG-VQVTESGKFHISPEGFLTINDVGPADAGRYECVARNTIGSASVSMVLSVNV----- 630

Qy        390 WSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVI 435
                                 ||    |   ||  |:::  |:| :
Db        631 --------------------PD---VSRNGDPFVATSIVEAIATV 652


RESULT 58
US-10-331-496A-28
; Sequence 28, Application US/10331496A
; Publication No. US20030228305A1
; GENERAL INFORMATION:
;  APPLICANT: FRANTZ,GRETCHEN
;  APPLICANT:  HILLAN,KENNETH J.
;  APPLICANT:  PHILLIPS,HEIDI S.
;  APPLICANT:  POLAKIS,PAUL
;  APPLICANT:  SMITH,VICTORIA
;  APPLICANT:  SPENCER,SUSAN D.
;  APPLICANT:  WILLIAMS,P. MICKEY
;  APPLICANT:  WU,THOMAS D.
;  APPLICANT:  ZHANG,ZEMIN
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF TUMOR
;  FILE REFERENCE: P5014R1-PCT
;  CURRENT APPLICATION NUMBER: US/10/331,496A
;  CURRENT FILING DATE:  2002-12-30
;  PRIOR APPLICATION NUMBER: US 60/345,444
;  PRIOR FILING DATE: 2002-01-02
;  PRIOR APPLICATION NUMBER: US 60/351,885
;  PRIOR FILING DATE: 2002-01-25
;  PRIOR APPLICATION NUMBER: US 60/360,066
;  PRIOR FILING DATE: 2002-02-25
;  PRIOR APPLICATION NUMBER: US 60/362,004
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/366,869
;  PRIOR FILING DATE: 2002-03-20
;  PRIOR APPLICATION NUMBER: US 60/366,284
;  PRIOR FILING DATE: 2002-03-21
;  PRIOR APPLICATION NUMBER: US 60/368,679
;  PRIOR FILING DATE: 2002-03-28
;  PRIOR APPLICATION NUMBER: US 60/404,809
;  PRIOR FILING DATE: 2002-08-19
;  PRIOR APPLICATION NUMBER: US 60/405,645
;  PRIOR FILING DATE: 2002-08-21
;  NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 28
;   LENGTH: 1496
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-331-496A-28

  Query Match             6.5%;  Score 152;  DB 15;  Length 1496;
  Best Local Similarity   21.5%;  Pred. No. 0.017;
  Matches  100;  Conservative   57;  Mismatches  155;  Indels  154;  Gaps   19;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |:||  || |      :  |  :|               ::|: : ||| :| | |
Db        277 GNTVYFTCRAEGNPKPEIIWLRNN---------------NELSMKTDSRLNLLDDGT-- 318

Qy         94 LIIKNLKIEDSDTYICEV-----EDQKEEVQLLVFGLTAN-------SDTHLLQGQSLTL 141
              |:|:| :  |   | |       | : :|| |  ||  |         :| :| |:|:||
Db        319 LMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTL 378

Qy        142 ----TLESPPGSSPSVQCRSP----RGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
                  |   ||  | :  |:|       ||     | :  :  ||| : |:   |   |
Db        379 ECSATGHPPPRISWTRGDRTPLPVDPRVNITPSGGLYIQNVVQGDSGEYACSATNNIDSV 438

Qy        194 EFKIDIVVLAFQKASSIVYKK---EGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKS 250
                |:| |  : :    :   ||: |:|                   | |   :
Db        439 HATAFIIVQALPQFTVTPQDRVVIEGQTVDF-----------------QCEAKGNPPP 479

Qy        251 WITFDLKNKEVSVKR---VTQDPKLQM-------------------GKKLPLHLTL-PQ 286
               | :    ::|| |   |      |::                        :|:  |||: |:
Db        480 VIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPR 539

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLEN 346
               | :|  : |          ||  |       | |   |  | :        |
Db        540 VTPVFASIPSDTTV----------EVGANV--------QLPCSSQGEPEPAITW-----N 576

Qy        347 KEAKVSKREKPVWVLNPE------------AGMWQCLL-----SDSGQVLLESNIKVLPT 389
              |:  |   |   : ::||              || ::|:      | |  ::|  |:
Db        577 KDG-VQVTESGKFHISPEGFLTINDVGPADAGRYECVARNTIGSASVSMVLSVNV----- 630

Qy        390 WSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVI 435
                                 ||    |   ||  |:::  |:| :
Db        631 --------------------PD---VSRNGDPFVATSIVEAIATV 652


RESULT 59
US-10-276-774-1957
```

```
; Sequence 1957, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
;  APPLICANT: Hyseq, Inc.
;  APPLICANT:  Tang, Y, Tom et al
;  TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
;  FILE REFERENCE: 21272-030
;  CURRENT APPLICATION NUMBER: US/10/276,774
;  CURRENT FILING DATE:  2002-11-18
;  PRIOR APPLICATION NUMBER: 09/560,875
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  NUMBER OF SEQ ID NOS: 2700
;  SOFTWARE: Custom
; SEQ ID NO 1957
;   LENGTH: 1498
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-276-774-1957

  Query Match             6.5%;  Score 152;  DB 12;  Length 1498;
  Best Local Similarity   21.5%;  Pred. No. 0.017;
  Matches  100;  Conservative   57;  Mismatches  155;  Indels  154;  Gaps   19;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |:||  || |      :  |  :|             ::|: : ||| :| | |
Db        277 GNTVYFTCRAEGNPKPEIIWLRNN---------------NELSMKTDSRLNLLDDGT-- 318

Qy         94 LIIKNLKIEDSDTYICEV-----EDQKEEVQLLVFGLTAN-------SDTHLLQGQSLTL 141
              |:|:| :  |   | |       | : :|| |  ||  |         :| :| |:|:||
Db        319 LMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTL 378

Qy        142 ----TLESPPGSSPSVQCRSP----RGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
                  |  ||  | :  |:|       ||      | :  :  ||| : |:  |   |
Db        379 ECSATGHPPPRISWTRGDRTPLPVDPRVNITPSGGLYIQNVVQGDSGEYACSATNNIDSV 438

Qy        194 EFKIDIVVLAFQKASSIVYKK---EGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKS 250
                |:| |  : :    :   ||: |:|              | |   :
Db        439 HATAFIIVQALPQPTVTPQDRVVIEGQTVDF------------------QCEAKGNPPP 479

Qy        251 WITFDLKNKEVSVKR---VTQDPKLQM--------------------GKKLPLHLTL-PQ 286
               | :    ::|| |   |     |::                    :|:  |||: |:
Db        480 VIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPR 539

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLEN 346
               | :|   : |             ||   |       | |   |   | :       |
Db        540 VTPVFASIPSDTTV----------EVGANV--------QLPCSSQGEPEPAITW-----N 576

Qy        347 KEAKVSKREKPVWVLNPE------------AGMWQCLL-----SDSGQVLLESNIKVLPT 389
              |:  |   |   : ::||            || ::|:      | |  ::|  |:
Db        577 KDG-VQVTESGKFHISPEGFLTINDVGPADAGRYECVARNTIGSASVSMVLSVNV----- 630

Qy        390 WSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVI 435
                                 ||    |   ||  |:::  |:| :
Db        631 --------------------PD---VSRNGDPFVATSIVEAIATV 652


RESULT 60
US-10-243-552-899
; Sequence 899, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
;  APPLICANT: Tang, Y. Tom
;  APPLICANT:  Yang, Yonghong
;  APPLICANT:  Wang, Zhiwei
;  APPLICANT:  Weng, Gezhi
;  APPLICANT:  Ma, Yunqing
;  TITLE OF INVENTION: Novel Nucleic Acids and
;  TITLE OF INVENTION:  Polypeptides
;  FILE REFERENCE: 807A
;  CURRENT APPLICATION NUMBER: US/10/243,552
;  CURRENT FILING DATE:  2002-09-12
;  PRIOR APPLICATION NUMBER: US 60/322,511
;  PRIOR FILING DATE: 2001-09-13
;  PRIOR APPLICATION NUMBER: PCT/US00/35017
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: US 09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: PCT/US01/02623
;  PRIOR FILING DATE: 2001-01-25
;  PRIOR APPLICATION NUMBER: US 09/491,404
;  PRIOR FILING DATE: 2000-01-25
;  PRIOR APPLICATION NUMBER: PCT/US01/03800
;  PRIOR FILING DATE: 2001-02-05
;  PRIOR APPLICATION NUMBER: US 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  PRIOR APPLICATION NUMBER: US 09/560,875
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: PCT/US01/04927
;  PRIOR FILING DATE: 2001-02-26
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 998
;  SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 899
;   LENGTH: 1498
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-243-552-899

  Query Match             6.5%;  Score 152;  DB 12;  Length 1498;
  Best Local Similarity   21.5%;  Pred. No. 0.017;
  Matches  100;  Conservative   57;  Mismatches  155;  Indels  154;  Gaps   19;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |:||  || |      :  |  :|             ::|: : ||| :| | |
Db        277 GNTVYFTCRAEGNPKPEIIWLRNN---------------NELSMKTDSRLNLLDDGT-- 318

Qy         94 LIIKNLKIEDSDTYICEV-----EDQKEEVQLLVFGLTAN-------SDTHLLQGQSLTL 141
              |:|:| :  |   | |       | : :|| |  ||  |         :| :| |:|:||
Db        319 LMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTL 378

Qy        142 ----TLESPPGSSPSVQCRSP----RGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
                  |  ||  | :  |:|       ||      | :  :  ||| : |:  |   |
Db        379 ECSATGHPPPRISWTRGDRTPLPVDPRVNITPSGGLYIQNVVQGDSGEYACSATNNIDSV 438

Qy        194 EFKIDIVVLAFQKASSIVYKK---EGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKS 250
                |:| |  : :    :   ||: |:|              | |   :
Db        439 HATAFIIVQALPQPTVTPQDRVVIEGQTVDF------------------QCEAKGNPPP 479

Qy        251 WITFDLKNKEVSVKR---VTQDPKLQM--------------------GKKLPLHLTL-PQ 286
               | :    ::|| |   |     |::                    :|:  |||: |:
Db        480 VIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPR 539

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLEN 346
               | :|   : |             ||   |       | |   |   | :       |
Db        540 VTPVFASIPSDTTV----------EVGANV--------QLPCSSQGEPEPAITW-----N 576

Qy        347 KEAKVSKREKPVWVLNPE------------AGMWQCLL-----SDSGQVLLESNIKVLPT 389
              |:  |   |   : ::||            || ::|:      | |  ::|  |:
Db        577 KDG-VQVTESGKFHISPEGFLTINDVGPADAGRYECVARNTIGSASVSMVLSVNV----- 630

Qy        390 WSTPVHPRASALPAPPTGSALPDPQTASALPDPPAASALPAALAVI 435
                                 ||    |   ||  |:::  |:| :
Db        631 --------------------PD---VSRNGDPFVATSIVEAIATV 652


RESULT 61
US-10-076-674-4
; Sequence 4, Application US/10076674
; Publication No. US20030165478A1
; GENERAL INFORMATION:
;  APPLICANT: Sokoll, Kenneth K.
;  TITLE OF INVENTION: Stabilized Synthetic Immunogen Delivery System
```

```
;  FILE REFERENCE: Immunogen Delivery System
;  CURRENT APPLICATION NUMBER: US/10/076,674
;  CURRENT FILING DATE:  2002-04-23
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 50
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (20)..(20)
;   OTHER INFORMATION: Xaa indicates epsilon-Lys
US-10-076-674-4

  Query Match             6.5%;  Score 150;  DB 14;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 0.00023;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         64 NQGSFLTKGPSKLNDRADSRRSLWDQGN 91
              ||||||||||||||||||||||||||||
Db         22 NQGSFLTKGPSKLNDRADSRRSLWDQGN 49


RESULT 62
US-10-076-674-5
; Sequence 5, Application US/10076674
; Publication No. US20030165478A1
; GENERAL INFORMATION:
;  APPLICANT: Sokoll, Kenneth K.
;  TITLE OF INVENTION: Stabilized Synthetic Immunogen Delivery System
;  FILE REFERENCE: Immunogen Delivery System
;  CURRENT APPLICATION NUMBER: US/10/076,674
;  CURRENT FILING DATE:  2002-04-23
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 50
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (20)..(20)
;   OTHER INFORMATION: Xaa indicates epsilon-Lys
US-10-076-674-5

  Query Match             6.5%;  Score 150;  DB 14;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 0.00023;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         64 NQGSFLTKGPSKLNDRADSRRSLWDQGN 91
              ||||||||||||||||||||||||||||
Db         22 NQGSFLTKGPSKLNDRADSRRSLWDQGN 49


RESULT 63
US-10-355-161A-4
; Sequence 4, Application US/10355161A
; Publication No. US20040009897A1
; GENERAL INFORMATION:
;  APPLICANT: Sokoll, Kenneth K.
;  TITLE OF INVENTION: Stabilized Synthetic Immunogen Delivery System
;  FILE REFERENCE: Immunogen Delivery System
;  CURRENT APPLICATION NUMBER: US/10/355,161A
;  CURRENT FILING DATE:  2003-01-31
;  PRIOR APPLICATION NUMBER: US 10/076674
;  PRIOR FILING DATE: 2002-02-14
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 50
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (20)..(20)
;   OTHER INFORMATION: Xaa indicates epsilon-Lys
US-10-355-161A-4

  Query Match             6.5%;  Score 150;  DB 15;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 0.00023;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         64 NQGSFLTKGPSKLNDRADSRRSLWDQGN 91
              ||||||||||||||||||||||||||||
Db         22 NQGSFLTKGPSKLNDRADSRRSLWDQGN 49


RESULT 64
US-10-355-161A-5
; Sequence 5, Application US/10355161A
; Publication No. US20040009897A1
; GENERAL INFORMATION:
;  APPLICANT: Sokoll, Kenneth K.
;  TITLE OF INVENTION: Stabilized Synthetic Immunogen Delivery System
;  FILE REFERENCE: Immunogen Delivery System
;  CURRENT APPLICATION NUMBER: US/10/355,161A
;  CURRENT FILING DATE:  2003-01-31
;  PRIOR APPLICATION NUMBER: US 10/076674
;  PRIOR FILING DATE: 2002-02-14
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
;   LENGTH: 50
;   TYPE: PRT
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (20)..(20)
;   OTHER INFORMATION: Xaa indicates epsilon-Lys
US-10-355-161A-5

  Query Match             6.5%;  Score 150;  DB 15;  Length 50;
  Best Local Similarity   100.0%;  Pred. No. 0.00023;
  Matches   28;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         64 NQGSFLTKGPSKLNDRADSRRSLWDQGN 91
              ||||||||||||||||||||||||||||
Db         22 NQGSFLTKGPSKLNDRADSRRSLWDQGN 49


RESULT 65
US-10-274-583-20
; Sequence 20, Application US/10274583
; Publication No. US20030138431A1
; GENERAL INFORMATION:
;  APPLICANT: Exelixis, Inc.
;  TITLE OF INVENTION: LRRCAPs AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
;  FILE REFERENCE: EX02-119C
;  CURRENT APPLICATION NUMBER: US/10/274,583
;  CURRENT FILING DATE:  2002-10-21
;  PRIOR APPLICATION NUMBER: 60/338,733
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/357,600
;  PRIOR FILING DATE: 2002-02-15
;  PRIOR APPLICATION NUMBER: 60/361,196
;  PRIOR FILING DATE: 2002-03-01
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
;   LENGTH: 1477
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-274-583-20

  Query Match             6.3%;  Score 146;  DB 14;  Length 1477;
```

```
  Best Local Similarity   22.2%;  Pred. No. 0.049;
  Matches   88;  Conservative   49;  Mismatches  140;  Indels  120;  Gaps   16;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |:||  || |      :  | :|               ::|: : ||| :| | |
Db        260 GNTVYFTCRAEGNPKPEIIWLRNN---------------NELSMKTDSRLNLLDDGT-- 301

Qy         94 LIIKNLKIEDSDTYICEV-----EDQKEEVQLLVFGLTAN-------SDTHLLQGQSLTL 141
              |:|:| :  |   | |       | : :|| |  ||  |         :| :| |:|:||
Db        302 LMIQNTQETDQGIYQCMAKNVAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTL 361

Qy        142 ----TLESPPGSSPSVQCRSP----RGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKV 193
                  |   ||  | :   |:|       ||      | :  :   ||| : |:   |  |
Db        362 ECSATGHPPPRISWTRGDRTPLPVDPRVNITPSGGLYIQNVVQGDSGEYACSATNNIDSV 421

Qy        194 EFKIDIVVLAFQKASSIVYKK---EGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKS 250
                |:| |  : :     :   ||: |:|                      | |   :
Db        422 HATAFIIVQALPQFTVTPQDRVVIEGQTVDF------------------QCEAKGNPPP 462

Qy        251 WITFDLKNKEVSVKR---VTQDPKLQM-------------------GKKLPLHLTL-PQ 286
                |  :     ::|| |   |     |::                    :|:  |||: |:
Db        463 VIAWTKGGSQLSVDRRHLVLSSGTLRISGVALHDQGQYECQAVNIIGSQKVVAHLTVQPR 522

Qy        287 ALPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLEN 346
                | :|   : |            ||   |          | |   |   | :       |
Db        523 VTPVFASIPSDTTV----------EVGANV--------QLPCSSQGEPEPAITW-----N 559

Qy        347 KEAKVSKREKPVWVLNPE------------AGMWQCL 371
              |:  |   |   : ::||             || ::|:
Db        560 KDG-VQVTESGKFHISPEGFLTINDVGPADAGRYECV 595


RESULT 66
US-10-168-417A-4
; Sequence 4, Application US/10168417A
; Publication No. US20040009185A1
; GENERAL INFORMATION:
;  APPLICANT: Emtage, Peter
;  APPLICANT:  Barber, Brian
;  APPLICANT:  Sambhara, Suryprakash
;  APPLICANT:  Sia, Charles Dwo Yuan
;  TITLE OF INVENTION: Enhancing the Immune Response to an Antigen by Presensitizing wit
;  TITLE OF INVENTION:  Inducing Agent Prior to Immunizing with the Inducing Agent and t
;  FILE REFERENCE: 11014-18-US
;  CURRENT APPLICATION NUMBER: US/10/168,417A
;  CURRENT FILING DATE:  2002-11-15
;  PRIOR APPLICATION NUMBER: US 60/174,587
;  PRIOR FILING DATE: 2000-01-05
;  PRIOR APPLICATION NUMBER: PCT/CA01/00005
;  PRIOR FILING DATE: 2001-01-05
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
;   LENGTH: 701
;   TYPE: PRT
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: modified CEA
US-10-168-417A-4

  Query Match             6.2%;  Score 145;  DB 15;  Length 701;
  Best Local Similarity   18.9%;  Pred. No. 0.021;
  Matches   96;  Conservative   75;  Mismatches  216;  Indels  122;  Gaps   19;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
              |: :  |   :  | | |  ||    : :    | |:  : :
Db        147 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV----------------- 187

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
                | |      ||  | : |:   |: :| || ::     : : | | :::|  | :
Db        188 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 245

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
```

```
              :|    |::| |:  :   |:|  |       |   | : | :  : : :||::||
Db        246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 303

Qy        189 NQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERASS 247
              :   :       :  ::     :       ||    :| | | ::  :  |||   ::
Db        304 SDTGLNRTTVTTITVYEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSLP 363

Qy        248 SKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQAL 288
                  :     |: :::  ||::           :| :    |: |         |:  :
Db        364 VSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPSY 423

Qy        289 PQYAGSGNLTLALEAKT-----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPTS 335
                |    ||:|:  | :           | : |    | :   |:    | ||:     |
Db        424 TYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSAS 483

Qy        336 PKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLES 382
                  ::|     |::      |   |||       :   |||      |   :||
Db        484 GHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ----- 538

Qy        383 NIKVLPTWSTPVHPRASALPAPPT---------------------GSALPDPQTASAL- 419
                    | || ||       |                       :   || |   |
Db        539 --------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLY 590

Qy        420 -PDPPAASALPAALAVISFLLGLGLGVAC 447
               || |  |   :     |:| |  | ::|
Db        591 GPDTPIISPPDS-----SYLSGADLNLSC 614


RESULT 67
US-10-176-847-32
; Sequence 32, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
;  APPLICANT: Veiby, Petter Ole
;  TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
;  TITLE OF INVENTION:  AND OVARIAN CANCER
;  FILE REFERENCE: MRI-039
;  CURRENT APPLICATION NUMBER: US/10/176,847
;  CURRENT FILING DATE:  2002-06-21
;  NUMBER OF SEQ ID NOS: 112
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 647
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-176-847-32

  Query Match             6.2%;  Score 143;  DB 14;  Length 647;
  Best Local Similarity   20.9%;  Pred. No. 0.027;
  Matches  105;  Conservative   81;  Mismatches  151;  Indels  166;  Gaps   25;

Qy         16 LALLPAATQGNKV-------VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILG---NQ 65
              | :: ||:|  |:        | : ||:| |||: :  :|  | |:      : |   |:
Db         15 LWIMFAASQAFKIETTPESRYLAQIGDSVSLTCSTTGCESPFFSWRTQIDSPLNGKVTNE 74

Qy         66 G--SFLTKGPSKLNDR--------ADSRRSLWDQG------NFP---------------- 93
              |  | ||  |    :           :||:   ::|       :||
Db         75 GTTSTLTMNPVSFGNEHSYLCTATCESRK--LEKGIQVEIYSFPKDPEIHLSGPLEAGKP 132

Qy         94 ---------------LIIKNLK----------IEDSD----------TYICEVED---- 113
                            | |  ||          :||:|           |:   :||
Db        133 ITVKCSVADVYPFDRLEIDLLKGDHLMKSQEFLEDADRKSLETKSLEVTFTPVIEDIGKV 192

Qy        114 -------------------QKEEVQLLVFG----LTANSDTHLLQGQSLTLTLESPPGS 149
                                :|:|: :      :: |  | | :| |:|:|  |
Db        193 LVCRAKLHIDEMDSVPTVRQAVKELQVYISPKNTVISVNPSTKLQEGGSVTMTCSSEGLP 252

Qy        150 SPSV----QCRSPRGKNIQGGKTLSVSQLELQDSGTWTC----TVLQNQKKVEFKIDIVV 201
              :| :    :  :   ::: |  ||::  : ::||| : |      : :|:|:||    ::|
Db        253 APEIFWSKKLDNGNLQHLSGNATLTLIAMRMEDSGIYVCEGVNLIGKNRKEVE----LIV 308

Qy        202 LAFQKASSIVYK---KEGEQVEFS------FPL-AFTVEKLTGSGELWWQAERASSSKSW 251
```

```
Db        309 QAFPRDPEIEMSGGLVNGSSVTVSCKVPSVYPLDRLEIELLKGETIL----------EN 357

Qy        252 ITF----DLKNKEVSVKRVTQDPKLQ-MGKKL----PLHLTLPQALPQYAGSGNLTLALE 302

Db        358 IEFLEDTDMKSLENKSLEMTFIPTIEDTGKALVCQAKLHIDDMEFEPKQRQS---TQTLY 414

Qy        303 AKTGKLHQEVNLVVMRATQLQK----NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPV 358

Db        415 VNVAP--RDTTVLVSPSSILEEGSSVNMTCLSQGFPAPKILWSRQLPNGELQPLSENATL 472

Qy        359 WVLNPEAGMWQCLLSDSGQVLLE 381

Db        473 TLISTK-------MEDSGVYLCE 488


RESULT 68
US-10-234-041-6
; Sequence 6, Application US/10234041
; Publication No. US20030153731A1
; GENERAL INFORMATION:
;  APPLICANT: Biogen, Inc.
;  APPLICANT:  Hession, Catherine A.
;  APPLICANT:  Lobb, Roy R.
;  APPLICANT:  Goelz, Susan E.
;  APPLICANT:  Osborn, Laurelee
;  APPLICANT:  Benjamin, Christopher D.
;  APPLICANT:  Rosa, Margaret D.
;  TITLE OF INVENTION: Endothelial Cell-Leukocyte Adhesion
;  TITLE OF INVENTION:  Molecules (ELAMs) and Molecules Involved in Leukocyte
;  TITLE OF INVENTION:  Adhesion (MILAs)
;  FILE REFERENCE: B124CP2DV2CN
;  CURRENT APPLICATION NUMBER: US/10/234,041
;  CURRENT FILING DATE:  2003-04-08
;  PRIOR APPLICATION NUMBER: 08/473,764
;  PRIOR FILING DATE: 1995-06-07
;  PRIOR APPLICATION NUMBER: 08/342,642
;  PRIOR FILING DATE: 1994-11-21
;  PRIOR APPLICATION NUMBER: 07/608,298
;  PRIOR FILING DATE: 1990-10-31
;  PRIOR APPLICATION NUMBER: 07/452,675
;  PRIOR FILING DATE: 1989-12-18
;  PRIOR APPLICATION NUMBER: 07/359,516
;  PRIOR FILING DATE: 1989-06-01
;  PRIOR APPLICATION NUMBER: 07/345,151
;  PRIOR FILING DATE: 1989-04-28
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 647
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-234-041-6

  Query Match             6.2%;  Score 143;  DB 14;  Length 647;
  Best Local Similarity   20.9%;  Pred. No. 0.027;
  Matches  105;  Conservative   81;  Mismatches  151;  Indels  166;  Gaps   25;

Qy         16 LALLPAATQGNKV-------VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILG---NQ 65

Db         15 LWIMFAASQAFKIETTPESRYLAQIGDSVSLTCSTTGCESPFFSWRTQIDSPLNGKVTNE 74

Qy         66 G--SFLTKGPSKLNDR--------ADSRRSLWDQG------NFP--------------- 93

Db         75 GTTSTLTMNPVSFGNEHSYLCTATCESRK--LEKGIQVEIYSFPKDPEIHLSGPLEAGKP 132

Qy         94 --------------LIIKNLK----------IEDSD----------TYICEVED---- 113

Db        133 ITVKCSVADVYPFDRLEIDLLKGDHLMKSQEFLEDADRKSLETKSLEVTFTPVIEDIGKV 192

Qy        114 -------------------QKEEVQLLVFG----LTANSDTHLLQGQSLTLTLESPPGS 149

Db        193 LVCRAKLHIDEMDSVPTVRQAVKELQVYISPKNTVISVNPSTKLQEGGSVTMTCSSEGLP 252

Qy        150 SPSV----QCRSPRGKNIQGGKTLSVSQLELQDSGTWTC----TVLQNQKKVEFKIDIVV 201

Db        253 APEIFWSKKLDNGNLQHLSGNATLTLIAMRMEDSGIYVCEGVNLIGKNRKEVE----LIV 308

Qy        202 LAFQKASSIVYK---KEGEQVEFS------FPL-AFTVEKLTGSGELWWQAERASSSKSW 251

Db        309 QAFPRDPEIEMSGGLVNGSSVTVSCKVPSVYPLDRLEIELLKGETIL----------EN 357

Qy        252 ITF----DLKNKEVSVKRVTQDPKLQ-MGKKL----PLHLTLPQALPQYAGSGNLTLALE 302

Db        358 IEFLEDTDMKSLENKSLEMTFIPTIEDTGKALVCQAKLHIDDMEFEPKQRQS---TQTLY 414

Qy        303 AKTGKLHQEVNLVVMRATQLQK----NLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPV 358

Db        415 VNVAP--RDTTVLVSPSSILEEGSSVNMTCLSQGFPAPKILWSRQLPNGELQPLSENATL 472

Qy        359 WVLNPEAGMWQCLLSDSGQVLLE 381

Db        473 TLISTK-------MEDSGVYLCE 488


RESULT 69
US-10-032-189-128
; Sequence 128, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook II, John P
;  APPLICANT:  Tchernev, Velizar T
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Spytek, Kimberly A
;  APPLICANT:  Zerhusen, Bryan D
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Grosse, William M
;  APPLICANT:  Lepley, Denise M
;  APPLICANT:  Burgess, Catherine E
;  APPLICANT:  Shimkets, Richard A
;  APPLICANT:  Grosse, William M
;  APPLICANT:  Szekeres, Edward S
;  APPLICANT:  Vernet, Corine A.M.
;  APPLICANT:  Li, Li
;  APPLICANT:  Casman, Stacie J
;  APPLICANT:  Boldog, Ferenc L
;  APPLICANT:  Gorman, Linda
;  APPLICANT:  Gangolli, Esha A
;  APPLICANT:  Fernandes, Elma R
;  APPLICANT:  Rieger, Daniel K
;  APPLICANT:  Edinger, Shlomit R
;  APPLICANT:  Gunther, Erik
;  APPLICANT:  Millet, Isabelle
;  APPLICANT:  Sciore, Paul
;  APPLICANT:  Ellerman, Karen
;  APPLICANT:  MacDougall, John R
;  APPLICANT:  Smithson, Glennda
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-228
;  CURRENT APPLICATION NUMBER: US/10/032,189
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/257,495
;  PRIOR FILING DATE: 2000-12-21
;  PRIOR APPLICATION NUMBER: 60/258,171
;  PRIOR FILING DATE: 2000-12-20
;  PRIOR APPLICATION NUMBER: 60/269,940
;  PRIOR FILING DATE: 2001-02-20
;  PRIOR APPLICATION NUMBER: 60/274,192
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/277,826
;  PRIOR FILING DATE: 2001-03-22
;  PRIOR APPLICATION NUMBER: 60/279,840
;  PRIOR FILING DATE: 2001-03-29
;  PRIOR APPLICATION NUMBER: 60/282,981
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: 60/283,656
```

```
;  PRIOR FILING DATE: 2001-04-13
;  PRIOR APPLICATION NUMBER: 60/309,247
;  PRIOR FILING DATE: 2001-07-31
;  PRIOR APPLICATION NUMBER: 60/311,754
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 60/313,331
;  PRIOR FILING DATE: 2001-08-17
;  NUMBER OF SEQ ID NOS: 260
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 128
;   LENGTH: 5636
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (3003)
;   OTHER INFORMATION: Wherein Xaa is any amino acid.
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (3041)
;   OTHER INFORMATION: Wherein Xaa is any amino acid.
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (3367)
;   OTHER INFORMATION: Wherein Xaa is any amino acid.
US-10-032-189-128

  Query Match             6.2%;  Score 143;  DB 14;  Length 5636;
  Best Local Similarity   20.4%;  Pred. No. 0.53;
  Matches   93;  Conservative   56;  Mismatches  168;  Indels  138;  Gaps   21;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |  | | |             |:  : : |      |      |:|   |
Db        807 GSNVTLPCYVQGYPEPTIKWRRLDNMPIFSR--PFSVSSISQLRTGA------------ 851

Qy         94 LIIKNLKIEDSDTYICEVEDQ----KEEVQLLVFGLTA------NSDTHLLQGQSLTLTL 143
              | | ||   |  ||||| |:|    : |  : | || |       |  ::::|| |||
Db        852 LFILNLWASDKGTYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPC 911

Qy        144 ESPPGS---------------SPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTC--- 184
                |:               :| :  ||          :| : :::||| | :||
Db        912 TLLAGNPIPERRWIKNSAMLLQNPYITVRS--------DGSLHIERVQLQDGGEYTCVAS 963

Qy        185 TVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAER 244
              |     |    :  |:     |:   ||  :  : |                 :
Db        964 NVAGTNNKTTSVVVHVLPTIQHGQQILSTIEG--IPVTLP----------------CKA 1004

Qy        245 ASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK 304
              : : |  : :  | : :|    |   |   |    |:  |:       ||
Db       1005 SGNPKPSVIWSKKGELIS----TSSAKFSAGADGSLYVVSPEG----EESGEYVCTATNT 1056

Qy        305 TGKLHQEVNLVV---------MRATQLQK-----------NLTCEV---------W--- 331
              |   ::| | |          :|     |           | |||          |
Db       1057 AGYAKRKVQLTVYVRPRVFGDLRGLSQDKPVEISVLAGEEVTLPCEVKSLPPPIITWAKE 1116

Qy        332 ----GPTSPK--LMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSD-SGQV--LLES 382
                  | ||:   : |  ::  | : |           ::||: |: :: :| |   ::
Db       1117 TQLISPFSPRHTFLPSGSMKITETRTS----------DSGMYLCVATNIAGNVTQAVKL 1165

Qy        383 NIKVLPTWST-PVHPRASA-----LPAPPTGSALP 411
              |: | |      | | :        :|    |: ||
Db       1166 NVHVPPKIQRGPKHLKVQVGQRVDIPCNAQGTPLP 1200


RESULT 70
US-10-120-801-72
; Sequence 72, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
;  APPLICANT: Pena, Carol
;  APPLICANT:  Guo, Xiaojia
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Kekuda, Ramesh
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Mehraban, Fuad
;  APPLICANT:  Topper, James N.
;  APPLICANT:  Malyankar, Uriel
;  APPLICANT:  Wasserman, Scott
;  APPLICANT:  Edinger, Shlomit
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Gunther, Erik
;  APPLICANT:  Komuves, Laszlo
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-340
;  CURRENT APPLICATION NUMBER: US/10/120,801
;  CURRENT FILING DATE:  2002-04-11
;  PRIOR APPLICATION NUMBER: 60/285748
;  PRIOR FILING DATE: 2001-04-23
;  PRIOR APPLICATION NUMBER: 60/286068
;  PRIOR FILING DATE: 2001-04-24
;  PRIOR APPLICATION NUMBER: 60/286292
;  PRIOR FILING DATE: 2001-04-25
;  PRIOR APPLICATION NUMBER: 60/288334
;  PRIOR FILING DATE: 2001-05-03
;  PRIOR APPLICATION NUMBER: 60/291241
;  PRIOR FILING DATE: 2001-05-16
;  PRIOR APPLICATION NUMBER: 60/322284
;  PRIOR FILING DATE: 2001-09-14
;  PRIOR APPLICATION NUMBER: 60/285609
;  PRIOR FILING DATE: 2001-04-20
;  NUMBER OF SEQ ID NOS: 155
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72
;   LENGTH: 5636
;   TYPE: PRT
;   ORGANISM: human
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (3003)
;   OTHER INFORMATION: Wherein Xaa is any amino acid.
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (3041)
;   OTHER INFORMATION: Wherein Xaa is any amino acid.
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (3367)
;   OTHER INFORMATION: Wherein Xaa is any amino acid.
US-10-120-801-72

  Query Match             6.2%;  Score 143;  DB 15;  Length 5636;
  Best Local Similarity   20.4%;  Pred. No. 0.53;
  Matches   93;  Conservative   56;  Mismatches  168;  Indels  138;  Gaps   21;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |  | | |             |:  : : |      |      |:|   |
Db        807 GSNVTLPCYVQGYPEPTIKWRRLDNMPIFSR--PFSVSSISQLRTGA------------ 851

Qy         94 LIIKNLKIEDSDTYICEVEDQ----KEEVQLLVFGLTA------NSDTHLLQGQSLTLTL 143
              | | ||   |  ||||| |:|    : |  : | || |       |  ::::|| |||
Db        852 LFILNLWASDKGTYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPC 911

Qy        144 ESPPGS---------------SPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTC--- 184
                |:               :| :  ||          :| : :::||| | :||
Db        912 TLLAGNPIPERRWIKNSAMLLQNPYITVRS--------DGSLHIERVQLQDGGEYTCVAS 963

Qy        185 TVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAER 244
              |     |    :  |:     |:   ||  :  : |                 :
Db        964 NVAGTNNKTTSVVVHVLPTIQHGQQILSTIEG--IPVTLP----------------CKA 1004

Qy        245 ASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK 304
              : : |  : :  | : :|    |   |   |    |:  |:       ||
Db       1005 SGNPKPSVIWSKKGELIS----TSSAKFSAGADGSLYVVSPEG----EESGEYVCTATNT 1056

Qy        305 TGKLHQEVNLVV---------MRATQLQK-----------NLTCEV---------W--- 331
```

```
                |   ::| | |         :|     |             | |||         |
Db       1057 AGYAKRKVQLTVYVRPRVFGDLRGLSQDKPVEISVLAGEEVTLPCEVKSLPPPIITWAKE 1116

Qy        332 ----GPTSPK--LMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSD-SGQV--LLES 382
                  | ||:   : |  ::  | : |           ::||: |: :: :| |   ::
Db       1117 TQLISPFSPRHTFLPSGSMKITETRTS----------DSGMYLCVATNIAGNVTQAVKL 1165

Qy        383 NIKVLPTWST-PVHPRASA-----LPAPPTGSALP 411
                |: | |     | | :        :|    |: ||
Db       1166 NVHVPPKIQRGPKHLKVQVGQRVDIPCNAQGTPLP 1200


RESULT 71
US-10-023-634-93
; Sequence 93, Application US/10023634
; Publication No. US20030236389A1
; GENERAL INFORMATION:
;  APPLICANT: Shimkets, Richard A
;  APPLICANT:  Colman, Steven D
;  APPLICANT:  Spytek, Kimberly A
;  APPLICANT:  Ballinger, Robert A
;  APPLICANT:  Guo, Xiaojia
;  APPLICANT:  Tchernev, Velizar T
;  APPLICANT:  Shenoy, Suresh G
;  APPLICANT:  Li, Li
;  APPLICANT:  Ellerman, Karen
;  APPLICANT:  Zerhusen, Bryan D
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Casman, Stacie J
;  APPLICANT:  Boldog, Ferenc
;  APPLICANT:  Gusev, Vladimir Y
;  APPLICANT:  Burgess, Catherine E
;  APPLICANT:  Edinger, Shlomit R
;  APPLICANT:  Gangolli, Esha A
;  APPLICANT:  Malyankar, Uriel M
;  APPLICANT:  Gunther, Erik
;  APPLICANT:  Smithson, Glennda
;  APPLICANT:  Millet, Isabelle
;  APPLICANT:  Gerlach, Valerie
;  TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
;  TITLE OF INVENTION:  Using the Same
;  FILE REFERENCE: 21402-221
;  CURRENT APPLICATION NUMBER: US/10/023,634
;  CURRENT FILING DATE:  2002-06-28
;  PRIOR APPLICATION NUMBER: 60/256,025
;  PRIOR FILING DATE: 2000-12-15
;  PRIOR APPLICATION NUMBER: 60/265,163
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: 60/272,929
;  PRIOR FILING DATE: 2001-03-02
;  PRIOR APPLICATION NUMBER: 60/274,864
;  PRIOR FILING DATE: 2001-03-09
;  PRIOR APPLICATION NUMBER: 60/276,688
;  PRIOR FILING DATE: 2001-03-16
;  PRIOR APPLICATION NUMBER: 60/277,880
;  PRIOR FILING DATE: 2001-03-22
;  PRIOR APPLICATION NUMBER: 60/286,409
;  PRIOR FILING DATE: 2001-04-25
;  PRIOR APPLICATION NUMBER: 60/309,246
;  PRIOR FILING DATE: 2001-07-31
;  PRIOR APPLICATION NUMBER: 60/315,600
;  PRIOR FILING DATE: 2001-08-29
;  NUMBER OF SEQ ID NOS: 132
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 93
;   LENGTH: 5636
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (3003)
;   OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
;   OTHER INFORMATION: specification.
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (3041)
;   OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
;   OTHER INFORMATION: specification.
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (3367)
;   OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
;   OTHER INFORMATION: specification.
US-10-023-634-93

  Query Match             6.2%;  Score 143;  DB 15;  Length 5636;
  Best Local Similarity   20.4%;  Pred. No. 0.53;
  Matches   93;  Conservative   56;  Mismatches  168;  Indels  138;  Gaps   21;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
              |  | | |            |:  : : |       |      |:|   |
Db        807 GSNVTLPCYVQGYPEPTIKWRRLDNMPIFSR--PFSVSSISQLRTGA------------ 851

Qy         94 LIIKNLKIEDSDTYICEVEDQ----KEEVQLLVFGLTA------NSDTHLLQGQSLTLTL 143
              | | ||    |  ||||| |:|    : |  : | || |       |  ::::|| |||
Db        852 LFILNLWASDKGTYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPC 911

Qy        144 ESPPGS---------------SPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTC--- 184
                |:                :| :  ||          :| : :::||| | :||
Db        912 TLLAGNPIPERRWIKNSAMLLQNPYITVRS--------DGSLHIERVQLQDGGEYTCVAS 963

Qy        185 TVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAER 244
              |     |    :  |:   |    |:   ||  :  : |                   :
Db        964 NVAGTNNKTTSVVVHVLPTIQHGQQILSTIEG--IPVTLP----------------CKA 1004

Qy        245 ASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK 304
              : : |  : :  | : :|    |   |   |    |::  |:       ||
Db       1005 SGNPKPSVIWSKKGELIS----TSSAKFSAGADGSLYVVSPEG----EESGEYVCTATNT 1056

Qy        305 TGKLHQEVNLVV---------MRATQLQK-----------NLTCEV---------W--- 331
              |   ::| | |          :|     |                | |||         |
Db       1057 AGYAKRKVQLTVYVRPRVFGDLRGLSQDKPVEISVLAGEEVTLPCEVKSLPPPIITWAKE 1116

Qy        332 ----GPTSPK--LMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSD-SGQV--LLES 382
                  | ||:   : |  ::  | : |           ::||: |: :: :| |   ::
Db       1117 TQLISPFSPRHTFLPSGSMKITETRTS----------DSGMYLCVATNIAGNVTQAVKL 1165

Qy        383 NIKVLPTWST-PVHPRASA-----LPAPPTGSALP 411
                |: | |     | | :        :|    |: ||
Db       1166 NVHVPPKIQRGPKHLKVQVGQRVDIPCNAQGTPLP 1200


RESULT 72
US-10-408-765A-1895
; Sequence 1895, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
;  APPLICANT: Ghosh, Soumitra S.
;  APPLICANT:  Fahy, Eoin D.
;  APPLICANT:  Zhang, Bing
;  APPLICANT:  Gibson, Bradford W.
;  APPLICANT:  Taylor, Steven W.
;  APPLICANT:  Glenn, Gary M.
;  APPLICANT:  Warnock, Dale E.
;  TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
;  TITLE OF INVENTION:  IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
;  FILE REFERENCE: 660088.465
;  CURRENT APPLICATION NUMBER: US/10/408,765A
;  CURRENT FILING DATE:  2003-04-04
;  NUMBER OF SEQ ID NOS: 3077
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1895
;   LENGTH: 5636
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
```

```
;   NAME/KEY: VARIANT
;   LOCATION: 3003, 3041, 3367
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-10-408-765A-1895

  Query Match             6.2%;  Score 143;  DB 16;  Length 5636;
  Best Local Similarity   20.4%;  Pred. No. 0.53;
  Matches   93;  Conservative   56;  Mismatches  168;  Indels  138;  Gaps   21;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
Db        807 GSNVTLPCYVQGYPEPTIKWRRLDNMPIFSR--PFSVSSISQLRTGA------------- 851

Qy         94 LIIKNLKIEDSDTYICEVEDQ----KEEVQLLVFGLTA------NSDTHLLQGQSLTLTL 143
Db        852 LFILNLWASDKGTYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPC 911

Qy        144 ESPPGS----------------SPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTC--- 184
Db        912 TLLAGNPIPERRWIKNSAMLLQNPYITVRS--------DGSLHIERVQLQDGGEYTCVAS 963

Qy        185 TVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAER 244
Db        964 NVAGTNNKTTSVVVHVLPTIQHGQQILSTIEG--IPVTLP----------------CKA 1004

Qy        245 ASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK 304
Db       1005 SGNPKPSVIWSKKGELIS----TSSAKFSAGADGSLYVVSPEG----EESGEYVCTATNT 1056

Qy        305 TGKLHQEVNLVV---------MRATQLQK-----------NLTCEV---------W--- 331
Db       1057 AGYAKRKVQLTVYVRPRVFGDLRGLSQDKPVEISVLAGEEVTLPCEVKSLPPPIITWAKE 1116

Qy        332 ----GPTSPK--LMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSD-SGQV--LLES 382
Db       1117 TQLISPFSPRHTFLPSGSMKITETRTS----------DSGMYLCVATNIAGNVTQAVKL 1165

Qy        383 NIKVLPTWST-PVHPRASA-----LPAPPTGSALP 411
Db       1166 NVHVPPKIQRGPKHLKVQVGQRVDIPCNAQGTPLP 1200


RESULT 73
US-10-108-260A-3037
; Sequence 3037, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3037
;   LENGTH: 956
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-108-260A-3037

  Query Match             6.1%;  Score 141;  DB 15;  Length 956;
  Best Local Similarity   25.6%;  Pred. No. 0.066;
  Matches   83;  Conservative   59;  Mismatches  128;  Indels   54;  Gaps   15;

Qy         96 IKNLKIEDSDTYICEVEDQKEEVQLLVFGLTA--NSDTHLLQ--GQSLTLTLESPPGSSP 151
Db        592 VENLRL----LRACFEETKKEEIKEVPFETLAQWNLEHATLNEAGNFLVEVSNDVVGSSI 647

Qy        152 SVQCRSPRGKNIQGGKTLSVSQLEL--------QDSGTW--TCTVLQNQKK--VEFKIDI 199
Db        648 SKELRR---LNKRWRKLVSKTQLEMNLPLMIKKQDQPTFDNSGNILSKEEKATVEFSTDM 704

Qy        200 VVLAFQKASSIV-----YKKEGEQVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWITF 254
Db        705 SVELPENYNQNIKAGEKHEKENEEFTGQLKVAKDVEKLIGQVEI-WEAE----AKSVLDQ 759

Qy        255 D--LKNKEVSVKRV----TQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKL 308
Db        760 DDVDTSMEESLKHLIAKGSMFDELMARSEDMLQMDIQNISSQESFQHVLTTGLQAKIQEA 819

Qy        309 HQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNPEAGMW 368
Db        820 KEKVQINVVKLIAALKNLT-----DVSPDLDIRLKMEE-----SQKELESYMMRAQQLLG 869

Qy        369 QC-----LLSDSGQVLLESNIKVL 387
Db        870 QRESPGELISKHKEALIISNTKSL 893


RESULT 74
US-09-925-301-1133
; Sequence 1133, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1133
;   LENGTH: 737
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (7)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (140)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (186)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (194)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (308)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (534)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (535)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1133

  Query Match             6.0%;  Score 140.5;  DB 9;  Length 737;
  Best Local Similarity   18.7%;  Pred. No. 0.051;
  Matches   96;  Conservative   73;  Mismatches  215;  Indels  129;  Gaps   20;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
Db        182 PSISXNNSNPVEXK-DAVAFTCEPEVQNTTYLWWVNGQSLPV----------------- 222

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVE-----DQKEEVQLLVF----GLTAN-S 129
```

```
Db        223 --SPRLQLSNGNMTLTLLSVKRNDAGSYECEIQNPASANRSDPVTLNVLYGPDGPTISPS 280

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188

Db        281 KANYRPGENLNLSCHA--ASNPPAQYSWFXNGTFQQSTQELFIPNITVNNSGSYTCQAHN 338

Qy        189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246

Db        339 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 398

Qy        247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287

Db        399 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 458

Qy        288 LPQYAGSGNLTLALEAKT----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334

Db        459 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 518

Qy        335 S---------------PKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQV 378

Db        519 SGHSRTTVKTITVSAXXPKPSIS---SNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ- 574

Qy        379 LLESNIKVLPTWSTPVHPRASALPAPPT--------------------GSALPDPQTA 416

Db        575 -----------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTL 622

Qy        417 SAL--PDPPAASALPAALAVISFLLGLGLGVAC 447

Db        623 DVLYGPDTPIISPPDS-----SYLSGANLNLSC 650
```

RESULT 75
US-10-311-823-16
; Sequence 16, Application US/10311823
; Publication No. US20040116683A1
; GENERAL INFORMATION:
; APPLICANT: Bryan J. Boyle
; APPLICANT: Nancy Mize
; APPLICANT: Matthew Arterburn
; APPLICANT: Y. Tom Tang
; APPLICANT: George Yeung
; APPLICANT: Ping Zhou
; APPLICANT: Chenghua Liu
; APPLICANT: Vinod Asundi
; APPLICANT: Radoje T. Drmanac
; APPLICANT: Meng-Yun Wang
; APPLICANT: Lichuan Chen
; APPLICANT: Yea-Huey Yang
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NEUROTRIMIN-LIKE
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: HYS-17CIP/US
; CURRENT APPLICATION NUMBER: US/10/311,823
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US01/03651
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/632,085
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-311-823-16

Query Match 6.0%; Score 139.5; DB 16; Length 374;
Best Local Similarity 23.5%; Pred. No. 0.024;
Matches 82; Conservative 54; Mismatches 136; Indels 77; Gaps 18;

```
Qy         36 TVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI 95

Db         55 TVFLRCTVNSNPPARFIWKRGSD----------TLSHSQDNG-VDIYEPLYTQGETKVL 102

Qy         96 -IKNLKIEDSDTYICEVEDQKE---EVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSP 151

Db        103 KLKNLRPQDYASYTCQVSVRNVCGIPDKAITFRLTNTTAPPALK-LSVNETLVVNPGENV 161

Qy        152 SVQC----------------RSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN---- 189

Db        162 TVQCLLTGGDPLPQLQWSHGPGPLPLGALAQGG-TLSIPSVQARDSGYYNCTATNNVGNP 220

Qy        190 QKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFSFPL-------AFTVEKLTGSGELW 239

Db        221 AKKT---VNLLVRSMKNATFQITPDVIKESENIQLGQDLKLSCHVDAVPQEKVTYQ---W 274

Qy        240 WQ-AERASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHL----------TLPQA- 287

Db        275 FKNGKPARMSKRLL---VTRNDPELPAVTSSLEL-----IDLHFSDYGTYLCMASFPGAP 326

Qy        288 LPQYAGSGNLTLALEAKTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSP 336

Db        327 VPDLSVEVNISSETVPPTISVPKGRAVVTVREGS-PAELQCEVRGKPRP 374
```

RESULT 76
US-10-311-823-12
; Sequence 12, Application US/10311823
; Publication No. US20040116683A1
; GENERAL INFORMATION:
; APPLICANT: Bryan J. Boyle
; APPLICANT: Nancy Mize
; APPLICANT: Matthew Arterburn
; APPLICANT: Y. Tom Tang
; APPLICANT: George Yeung
; APPLICANT: Ping Zhou
; APPLICANT: Chenghua Liu
; APPLICANT: Vinod Asundi
; APPLICANT: Radoje T. Drmanac
; APPLICANT: Meng-Yun Wang
; APPLICANT: Lichuan Chen
; APPLICANT: Yea-Huey Yang
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NEUROTRIMIN-LIKE
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: HYS-17CIP/US
; CURRENT APPLICATION NUMBER: US/10/311,823
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US01/03651
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/632,085
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-311-823-12

Query Match 6.0%; Score 138.5; DB 16; Length 442;
Best Local Similarity 25.6%; Pred. No. 0.036;
Matches 60; Conservative 36; Mismatches 85; Indels 53; Gaps 11;

```
Qy         36 TVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI 95

Db        136 TVFLRCTVNSNPPARFIWKRGSD----------TLSHSQDNG-VDIYEPLYTQGETKVL 183

Qy         96 -IKNLKIEDSDTYICEVEDQKE---EVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSP 151
```

```
Db        184 KLKNLRPQDYASYTCQVSVRNVCGIPDKAITFRLTNTTAPPALK-LSVNETLVVNPGENV 242

Qy        152 SVQC-----------------RSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN---- 189
              :|||                 | |   ||| |||:  :: :||| : ||   |
Db        243 TVQCLLTGGDPLPQLQWSHGPGPLPLGALAQGG-TLSIPSVQARDSGYYNCTATNNVGNP 301

Qy        190 QKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFSFPL-------AFTVEKLT 233
               ||    ::::| : : |:  :     || | ::    |       |   ||:|
Db        302 AKKT---VNLLVRSMKNATFQITPDVIKESENIQLGQDLKLSCHVDAVPQEKVT 352


RESULT 77
US-10-311-823-4
; Sequence 4, Application US/10311823
; Publication No. US20040116683A1
; GENERAL INFORMATION:
;  APPLICANT: Bryan J. Boyle
;  APPLICANT:  Nancy Mize
;  APPLICANT:  Matthew Arterburn
;  APPLICANT:  Y. Tom Tang
;  APPLICANT:  George Yeung
;  APPLICANT:  Ping Zhou
;  APPLICANT:  Chenghua Liu
;  APPLICANT:  Vinod Asundi
;  APPLICANT:  Radoje T. Drmanac
;  APPLICANT:  Meng-Yun Wang
;  APPLICANT:  Lichuan Chen
;  APPLICANT:  Yea-Huey Yang
;  TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NEUROTRIMIN-LIKE
;  TITLE OF INVENTION:  POLYPEPTIDES AND POLYNUCLEOTIDES
;  FILE REFERENCE: HYS-17CIP/US
;  CURRENT APPLICATION NUMBER: US/10/311,823
;  CURRENT FILING DATE:  2003-09-29
;  PRIOR APPLICATION NUMBER: PCT/US01/03651
;  PRIOR FILING DATE: 2001-02-02
;  PRIOR APPLICATION NUMBER: 09/632,085
;  PRIOR FILING DATE: 2000-08-02
;  PRIOR APPLICATION NUMBER: 09/560,875
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 09/496,914
;  PRIOR FILING DATE: 2000-02-03
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 458
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-311-823-4

  Query Match             6.0%;  Score 138.5;  DB 16;  Length 458;
  Best Local Similarity   25.6%;  Pred. No. 0.038;
  Matches   60;  Conservative   36;  Mismatches   85;  Indels   53;  Gaps   11;

Qy         36 TVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFPLI 95
              || | || :     :| ||  :             |   |: |   |    |: ||   ::
Db        152 TVFLRCTVNSNPPARFIWKRGSD----------TLSHSQDNG-VDIYEPLYTQGETKVL 199

Qy         96 -IKNLKIEDSDTYICEVEDQKE---EVQLLVFGLTANSDTHLLQGQSLTLTLESPPGSSP 151
               :|||: :|  :| |:|  :       : : | ||  :    |:  |:  ||   || :
Db        200 KLKNLRPQDYASYTCQVSVRNVCGIPDKAITFRLTNTTAPPALK-LSVNETLVVNPGENV 258

Qy        152 SVQC-----------------RSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQN---- 189
              :|||                 | |   ||| |||:  :: :||| : ||   |
Db        259 TVQCLLTGGDPLPQLQWSHGPGPLPLGALAQGG-TLSIPSVQARDSGYYNCTATNNVGNP 317

Qy        190 QKKVEFKIDIVVLAFQKASSIVYK---KEGEQVEFSFPL-------AFTVEKLT 233
               ||    ::::| : : |:  :     || | ::    |       |   ||:|
Db        318 AKKT---VNLLVRSMKNATFQITPDVIKESENIQLGQDLKLSCHVDAVPQEKVT 368


RESULT 78
US-09-905-129-21
; Sequence 21, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
;  APPLICANT: Einat, et al
;  TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THERE
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 540579-2007.2
;  CURRENT APPLICATION NUMBER: US/09/905,129
;  CURRENT FILING DATE:  2001-07-13
;  PRIOR APPLICATION NUMBER: 09/802,318
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/207,821
;  PRIOR FILING DATE: 2000-05-30
;  PRIOR APPLICATION NUMBER: 60/084,944
;  PRIOR FILING DATE: 1998-05-11
;  PRIOR APPLICATION NUMBER: 60/085,673
;  PRIOR FILING DATE: 1998-05-15
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-905-129-21

  Query Match             5.9%;  Score 138;  DB 9;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy          10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
               | : | :| ||      |:  :   | :: : |||         |      :||
Db        2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy          68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
                        |    | | :  ||   ||   | |:||  :||  | |     |   :
Db        2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy         118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
               ||| |    ||:          | :  | :| |   :      | :  | |  : |
Db        2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy         165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
                     || |  :  :|:| : | | :|:      |  |:|:|:    | : | :|:
Db        2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy         215 EGE---------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
               | :                   :: :| |    |      |       :    :| ::
Db        2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy         254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
               |:   |   || ::                                |:  |   | : ||
Db        2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335

Qy         311 EVNLVVMRATQLQKNLT---------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
               | : |: |    :| |                || |   ||:   |   ||   |  :
Db        2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy         356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
                 ::             ::| : ||: :|       | :  |::
Db        2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ---------------- 2438

Qy         402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
                 ||  :  |:| |                |  :| |    | |  : :
Db        2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 79
US-09-991-630-21
; Sequence 21, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
;  APPLICANT: Einat, et al
;  TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THERE
;  TITLE OF INVENTION:  AND USES THEREOF
```

```
;  FILE REFERENCE: 540579-2007.3
;  CURRENT APPLICATION NUMBER: US/09/991,630
;  CURRENT FILING DATE:  2001-11-06
;  PRIOR APPLICATION NUMBER: 09/905,129
;  PRIOR FILING DATE: 2001-07-13
;  PRIOR APPLICATION NUMBER: 09/802,318
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 09/729,485
;  PRIOR FILING DATE: 2000-12-04
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-991-630-21

  Query Match             5.9%;  Score 138;  DB 9;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
              | : | :| ||      |:  :    | :: : |||         |       :||
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
                    |    | | :  ||   ||   | |:||  :||  | |     |   :
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
              ||| |    ||:          | :  | :| |   :      | :  | |  : |
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
                    || |   :  :|:| : | | :|:     |  |:|:|:    | : | :|:
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE---------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
              | :                    :: :| |     |      |       :    :| ::
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
              |:   |   || ::                          |: |   | : ||
Db       2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT--------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
              | : |: |    :| |              ||  |   ||:   |   ||    | :
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
               ::             ::| : ||: :|       | :  |::
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ--------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
               ||  :  |:| |                 | :| |   | |  : :
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 80
US-10-072-012-607
; Sequence 607, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
;  APPLICANT: Tchernev, Velizar
;  APPLICANT:  Spytek, Kimberly
;  APPLICANT:  Zerhusen, Bryan
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Shimkets, Richard
;  APPLICANT:  Li, Li
;  APPLICANT:  Gangolli, Esha
;  APPLICANT:  Padigaru, Muralidhara
;  APPLICANT:  Anderson, David W.
;  APPLICANT:  Rastelli, Luca
;  APPLICANT:  Miller, Charles E.
;  APPLICANT:  Gerlach, Valerie
;  APPLICANT:  Taupier Jr, Raymond J.
;  APPLICANT:  Gusev, Vladimir Y.
;  APPLICANT:  Colman, Steven D.
;  APPLICANT:  Wolenc, Adam R.
;  APPLICANT:  Pena, Carol E. A
;  APPLICANT:  Furtak, Katarzyna
;  APPLICANT:  Grosse, William M.
;  APPLICANT:  Alsobrook II, John P.
;  APPLICANT:  Lepley, Denise M.
;  APPLICANT:  Rieger, Daniel K.
;  APPLICANT:  Burgess, Catherine E.
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-258
;  CURRENT APPLICATION NUMBER: US/10/072,012
;  CURRENT FILING DATE:  2002-01-31
;  PRIOR APPLICATION NUMBER: 60/265,102
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: 60/265,514
;  PRIOR FILING DATE: 2001-01-31
;  PRIOR APPLICATION NUMBER: 60/265,517
;  PRIOR FILING DATE: 2001-01-31
;  PRIOR APPLICATION NUMBER: 60/265,412
;  PRIOR FILING DATE: 2001-01-31
;  PRIOR APPLICATION NUMBER: 60/265,395
;  PRIOR FILING DATE: 2001-01-31
;  PRIOR APPLICATION NUMBER: 60/266,406
;  PRIOR FILING DATE: 2001-02-02
;  PRIOR APPLICATION NUMBER: 60/266,767
;  PRIOR FILING DATE: 2001-02-05
;  PRIOR APPLICATION NUMBER: 60/267,057
;  PRIOR FILING DATE: 2001-02-07
;  PRIOR APPLICATION NUMBER: 60/266,975
;  PRIOR FILING DATE: 2001-02-07
;  PRIOR APPLICATION NUMBER: 60/267,459
;  PRIOR FILING DATE: 2001-02-08
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1391
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 607
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-072-012-607

  Query Match             5.9%;  Score 138;  DB 12;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
              | : | :| ||      |:  :    | :: : |||         |       :||
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
                    |    | | :  ||   ||   | |:||  :||  | |     |   :
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
              ||| |    ||:          | :  | :| |   :      | :  | |  : |
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
                    || |   :  :|:| : | | :|:     |  |:|:|:    | : | :|:
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE---------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
              | :                    :: :| |     |      |       :    :| ::
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
              |:   |   || ::                          |: |   | : ||
Db       2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335
```

```
Qy        311 EVNLVVMRATQLQKNLT---------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ---------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486
```

RESULT 81
US-10-072-012-608
; Sequence 608, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 608
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-072-012-608

Query Match 5.9%; Score 138; DB 12; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

```
Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSP 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE--------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
Db       2304 FNNGTLYFNEVGMRE---------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT---------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ---------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486
```

RESULT 82
US-10-454-351-21
; Sequence 21, Application US/10454351
; Publication No. US20040053301A1
; GENERAL INFORMATION:
; APPLICANT: Quark Biotech Inc.; Paz Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THERE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 010/PCT2-US2; EINAT=7E
; CURRENT APPLICATION NUMBER: US/10/454,351
; CURRENT FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: US 09/991,630
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 09/905,129
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: US 09/729,485
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/312,216
; PRIOR FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-454-351-21

Query Match 5.9%; Score 138; DB 12; Length 2828;

```
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
              | : | :| ||       |:  :    | :: : |||         |       :||
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
                        |    | | :  ||   ||   | |:||  :||  | |     |   :
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
              ||| |    ||:         | :  | :| |   :      | :  | |  : |
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
                    || |   :  :|:| : | | :|:      |  |:|:|:    | : | :|:
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE--------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
              | :                    :: :| |     |      |       :    :| ::
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
              |:    |   || ::                          |: |   | : ||
Db       2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT---------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
              | : |: |    :| |                || |   ||:   |   ||    | :
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
               ::             ::| : ||: :|       | :  |::
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ---------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
              ||  :  |:| |                  | :| |   | |  : :
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 83
US-09-802-318-21
; Sequence 21, Application US/09802318
; Publication No. US20020086825A1
; GENERAL INFORMATION:
;  APPLICANT: Einat, et al
;  TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS,EXPRESSION PRODUCTS THERE
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: 540579-2007.1
;  CURRENT APPLICATION NUMBER: US/09/802,318
;  CURRENT FILING DATE:  2001-03-08
;  PRIOR APPLICATION NUMBER: 09/632,862
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: 60/207,821
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-802-318-21

  Query Match             5.9%;  Score 138;  DB 12;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
              | : | :| ||       |:  :    | :: : |||         |       :||
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
                        |    | | :  ||   ||   | |:||  :||  | |     |   :
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
              ||| |    ||:         | :  | :| |   :      | :  | |  : |
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
                    || |   :  :|:| : | | :|:      |  |:|:|:    | : | :|:
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE--------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
              | :                    :: :| |     |      |       :    :| ::
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
              |:    |   || ::                          |: |   | : ||
Db       2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT---------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
              | : |: |    :| |                || |   ||:   |   ||    | :
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
               ::             ::| : ||: :|       | :  |::
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ---------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
              ||  :  |:| |                  | :| |   | |  : :
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 84
US-10-176-847-54
; Sequence 54, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
;  APPLICANT: Veiby, Petter Ole
;  TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
;  TITLE OF INVENTION:  AND OVARIAN CANCER
;  FILE REFERENCE: MRI-039
;  CURRENT APPLICATION NUMBER: US/10/176,847
;  CURRENT FILING DATE:  2002-06-21
;  NUMBER OF SEQ ID NOS: 112
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 54
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-176-847-54

  Query Match             5.9%;  Score 138;  DB 14;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
              | : | :| ||       |:  :    | :: : |||         |       :||
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
                        |    | | :  ||   ||   | |:||  :||  | |     |   :
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
              ||| |    ||:         | :  | :| |   :      | :  | |  : |
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
                    || |   :  :|:| : | | :|:      |  |:|:|:    | : | :|:
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE--------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
              | :                    :: :| |     |      |       :    :| ::
```

```
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
              |:   |   || ::                                 |: |   | : ||
Db       2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT--------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
              | : |: |    :| |                ||  |   ||:   |   ||    | :
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
                ::            ::| : ||: :|       | :  |::
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ--------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
                ||  :  |:| |                |  :| |    | |  : :
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 85
US-10-177-293-110
; Sequence 110, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Gannavarpu, Manjula
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Mertens, Maureen
;  APPLICANT:  Myer, Vic
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Hoersch, Sebastian
;  APPLICANT:  Monahan, John
;  APPLICANT:  Meyers, Rachel E.
;  APPLICANT:  Bast Jr., Robert C.
;  APPLICANT:  Hortobagyi, Gabriel N.
;  APPLICANT:  Pusztai, Lajos
;  APPLICANT:  Meric, Funda
;  APPLICANT:  Sahin, Aysegul
;  APPLICANT:  Mills, Gordon B.
;  TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
;  TITLE OF INVENTION:  PREVENTION, AND THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-038
;  CURRENT APPLICATION NUMBER: US/10/177,293
;  CURRENT FILING DATE:  2002-06-21
;  PRIOR APPLICATION NUMBER: US 60/299,887
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: US 60/301,572
;  PRIOR FILING DATE: 2001-06-27
;  PRIOR APPLICATION NUMBER: US 60/306,501
;  PRIOR FILING DATE: 2001-07-18
;  PRIOR APPLICATION NUMBER: US 60/325,002
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/362,585
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/xxx,xxx
;  PRIOR FILING DATE: 2002-05-14
;  NUMBER OF SEQ ID NOS: 506
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 110
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-177-293-110

  Query Match             5.9%;  Score 138;  DB 14;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
                | : | :| ||      |:  :    | :: : |||         |       :||

Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
                      |    | | :  ||   ||   | |:||  :||  | |      |   :
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
              ||| |    ||:          | :  | :| |   :      | :  | |  : |
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
                    || |  :  :|:| : | | :|:     |  |:|:|:    | : | :|:
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE---------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
              | :                   :: :| |    |     |       :    :| ::
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
              |:   |   || ::                                 |: |   | : ||
Db       2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT--------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
              | : |: |    :| |                ||  |   ||:   |   ||    | :
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
                ::            ::| : ||: :|       | :  |::
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ--------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
                ||  :  |:| |                |  :| |    | |  : :
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 86
US-10-301-822-49
; Sequence 49, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-10-301-822-49

  Query Match             5.9%;  Score 138;  DB 14;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
```

```
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117

Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164

Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214

Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE--------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253

Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310

Db       2304 FNNGTLYFNEVGMRE---------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT---------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355

Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401

Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ---------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434

Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 87
US-10-032-189-126
; Sequence 126, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook II, John P
;  APPLICANT:  Tchernev, Velizar T
;  APPLICANT:  Liu, Xiaohong
;  APPLICANT:  Spytek, Kimberly A
;  APPLICANT:  Zerhusen, Bryan D
;  APPLICANT:  Patturajan, Meera
;  APPLICANT:  Grosse, William M
;  APPLICANT:  Lepley, Denise M
;  APPLICANT:  Burgess, Catherine E
;  APPLICANT:  Shimkets, Richard A
;  APPLICANT:  Grosse, William M
;  APPLICANT:  Szekeres, Edward S
;  APPLICANT:  Vernet, Corine A.M.
;  APPLICANT:  Li, Li
;  APPLICANT:  Casman, Stacie J
;  APPLICANT:  Boldog, Ferenc L
;  APPLICANT:  Gorman, Linda
;  APPLICANT:  Gangolli, Esha A
;  APPLICANT:  Fernandes, Elma R
;  APPLICANT:  Rieger, Daniel K
;  APPLICANT:  Edinger, Shlomit R
;  APPLICANT:  Gunther, Erik
;  APPLICANT:  Millet, Isabelle
;  APPLICANT:  Sciore, Paul
;  APPLICANT:  Ellerman, Karen
;  APPLICANT:  MacDougall, John R
;  APPLICANT:  Smithson, Glennda
;  TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 21402-228
;  CURRENT APPLICATION NUMBER: US/10/032,189
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 60/257,495
;  PRIOR FILING DATE: 2000-12-21
;  PRIOR APPLICATION NUMBER: 60/258,171
;  PRIOR FILING DATE: 2000-12-20
;  PRIOR APPLICATION NUMBER: 60/269,940
;  PRIOR FILING DATE: 2001-02-20
;  PRIOR APPLICATION NUMBER: 60/274,192
;  PRIOR FILING DATE: 2001-03-08
;  PRIOR APPLICATION NUMBER: 60/277,826
;  PRIOR FILING DATE: 2001-03-22
;  PRIOR APPLICATION NUMBER: 60/279,840
;  PRIOR FILING DATE: 2001-03-29
;  PRIOR APPLICATION NUMBER: 60/282,981
;  PRIOR FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: 60/283,656
;  PRIOR FILING DATE: 2001-04-13
;  PRIOR APPLICATION NUMBER: 60/309,247
;  PRIOR FILING DATE: 2001-07-31
;  PRIOR APPLICATION NUMBER: 60/311,754
;  PRIOR FILING DATE: 2001-08-17
;  PRIOR APPLICATION NUMBER: 60/313,331
;  PRIOR FILING DATE: 2001-08-17
;  NUMBER OF SEQ ID NOS: 260
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 126
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-032-189-126

  Query Match             5.9%;  Score 138;  DB 14;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy          10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67

Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117

Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164

Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214

Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE--------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253

Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310

Db       2304 FNNGTLYFNEVGMRE---------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT---------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355

Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401

Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ---------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434

Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 88
US-10-295-027-58
; Sequence 58, Application US/10295027
```

```
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-295-027-58

  Query Match             5.9%;  Score 138;  DB 15;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE--------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
Db       2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT--------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ---------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 89
US-10-295-027-1175
; Sequence 1175, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1175
;   LENGTH: 2828
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-295-027-1175

  Query Match             5.9%;  Score 138;  DB 15;  Length 2828;
  Best Local Similarity   19.4%;  Pred. No. 0.5;
  Matches  103;  Conservative   63;  Mismatches  180;  Indels  184;  Gaps   21;

Qy         10 LLLVLQLALLPAATQGNKV--VLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGS 67
Db       2036 LAIRLHVAALPPVIHQEKLENISLPPGLSIHIHCTAKAAPLPSVRW-------VLG---- 2084

Qy         68 FLTKGPSKLNDRADSRRSLWDQGN---FP---LIIKNLKIEDSDTYICE----VEDQKEE 117
```

```
Db       2085 ----------DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPKDSGRYECVAANLVGSARRT 2134

Qy        118 VQLLVFGLTANS--------DTHLLQGQSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
              ||| |    ||:          | : | :| |   :       | : | |  : |
Db       2135 VQLNVQRAAANARITGTSPRRTDVRYGGTLKLDCSASGDPWPRILWRLPSKRMIDALFSF 2194

Qy        165 -------GGKTLSVSQLELQDSGTWTCTVLQNQ---KKVEFKIDIVVLAFQKASSIVYKK 214
                    || |   :  :|:| : | | :|:      |  |:|:|:    | : | :|:
Db       2195 DSRIKVFANGTLVVKSVTDKDAGDYLC-VARNKVGDDYVVLKVDVVM----KPAKIEHKE 2249

Qy        215 EGE--------------------QVEFSFPLAFTVEKLTGSGELWWQAERASSSKSWIT 253
              | :                    :: :| |    |      |       :    :| ::
Db       2250 ENDHKVFYGGDLKVDCVATGLPNPEISWSLPDGSLVNSFMQSD------DSGGRTKRYVV 2303

Qy        254 FD---LKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAKTGKLHQ 310
              |:   |   || ::                          |: |   | : ||
Db       2304 FNNGTLYFNEVGMRE--------------------------EGDYTCFAENQVGKDEM 2335

Qy        311 EVNLVVMRATQLQKNLT---------------CEVWGPTSPKLMLSLKLENKEAKVSKRE 355
              | .: |: |    :| |                ||  |   ||:   |   ||    |  :
Db       2336 RVRVKVVTAPATIRNKTYLAVQVPYGDVVTVACEAKGEPMPKVTW-LSPTNKVIPTSSEK 2394

Qy        356 KPVW---------VLNPEAGMWQCLLSDSG-----QVLLESNIKVLPTWSTPVHPRASAL 401
               ::              ::| : ||: :|       | :  |::
Db       2395 YQIYQDGTLLIQKAQRSDSGNYTCLVRNSAGEDRKTVWIHVNVQ---------------- 2438

Qy        402 PAPPTGSALPDPQT----------------ASALPDPPAASALPAALAV 434
               ||  :  |:| |                 | :| |   | |  : :
Db       2439 --PPKINGNPNPITTVREIAAGGSRKLIDCKAEGIPTPRVLWAFPEGVVL 2486


RESULT 90
US-10-253-286-282
; Sequence 282, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2015
;  CURRENT APPLICATION NUMBER: US/10/253,286
;  CURRENT FILING DATE:  2003-01-13
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 282
;   LENGTH: 702
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-253-286-282

  Query Match             5.9%;  Score 137.5;  DB 12;  Length 702;
  Best Local Similarity   18.8%;  Pred. No. 0.081;
  Matches   96;  Conservative   76;  Mismatches  215;  Indels  123;  Gaps   20;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
              |: :  |   :  | | |  ||    : :    | |:  : :
Db        147 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV----------------- 187

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
               | |      ||  | : |:   |: :| || ::     : : | | :::|  | :
Db        188 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 245

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
              :|    |::| |:  :   |:|  |       |   | : | :  : : :||::||
Db        246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 303

Qy        189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
              :  :   :  : :  :    :        ||    :| | | ::  :  |||   ::
Db        304 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 363

Qy        247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
                    :     |: :::  ||::           :| :    |: |         |:  :
Db        364 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 423

Qy        288 LPQYAGSGNLTLALEAKT----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
                  |    ||:|:  | :           | : |    | :   |:    | ||:
Db        424 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 483

Qy        335 SPKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLE 381
              |     ::|     |::      |   |||       :   |||     |   :||
Db        484 SGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ---- 539

Qy        382 SNIKVLPTWSTPVHPRASALPAPPT---------------------GSALPDPQTASAL 419
                    | || ||         |                      :   || |   |
Db        540 ---------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVL 590

Qy        420 --PDPPAASALPAALAVISFLLGLGLGVAC 447
                ||  |   |    :     |:| |   | ::|
Db        591 YGPDTPIISPPDS-----SYLSGANLNLSC 615


RESULT 91
US-10-380-136-16
; Sequence 16, Application US/10380136
; Publication No. US20040071726A1
; GENERAL INFORMATION:
;  APPLICANT: Chicz, Roman M.
;  APPLICANT:  Tomlinson, Andy
;  TITLE OF INVENTION: PEPTIDE EPITOPES RECOGNIZED BY ANTIGEN
;  TITLE OF INVENTION:  SPECIFIC CD4+ T LYMPHOCYTES
;  FILE REFERENCE: 08191-019US1
;  CURRENT APPLICATION NUMBER: US/10/380,136
;  CURRENT FILING DATE:  2003-03-11
;  PRIOR APPLICATION NUMBER: PCT/US01/28467
;  PRIOR FILING DATE: 2001-09-12
;  PRIOR APPLICATION NUMBER: US 60/232,185
;  PRIOR FILING DATE: 2000-09-12
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 702
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-380-136-16

  Query Match             5.9%;  Score 137.5;  DB 12;  Length 702;
  Best Local Similarity   18.8%;  Pred. No. 0.081;
  Matches   96;  Conservative   76;  Mismatches  215;  Indels  123;  Gaps   20;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
              |: :  |   :  | | |  ||    : :    | |:  : :
Db        147 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV----------------- 187

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
               | |      ||  | : |:   |: :| || ::     : : | | :::|  | :
Db        188 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 245

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
              :|    |::| |:  :   |:|  |       |   | : | :  : : :||::||
Db        246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 303

Qy        189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
              :  :   :  : :  :    :        ||    :| | | ::  :  |||   ::
Db        304 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 363

Qy        247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
                    :     |: :::  ||::           :| :    |: |         |:  :
Db        364 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 423

Qy        288 LPQYAGSGNLTLALEAKT----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
                  |    ||:|:  | :           | : |    | :   |:    | ||:
Db        424 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 483
```

```
Qy        335 SPKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLE 381
Db        484 SGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ---- 539

Qy        382 SNIKVLPTWSTPVHPRASALPAPPT----------------------GSALPDPQTASAL 419
Db        540 ---------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVL 590

Qy        420 --PDPPAASALPAALAVISFLLGLGLGVAC 447
Db        591 YGPDTPIISPPDS-----SYLSGANLNLSC 615


RESULT 92
US-10-157-031-341
; Sequence 341, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
;  APPLICANT: Baranova, A. V.
;  APPLICANT:  Yankovsky, N. K.
;  APPLICANT:  Kozlov, A. P.
;  APPLICANT:  Lobashev, A. V.
;  APPLICANT:  Krukovskaya, L. L.
;  TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
;  FILE REFERENCE: 2760-103
;  CURRENT APPLICATION NUMBER: US/10/157,031
;  CURRENT FILING DATE:  2002-05-30
;  NUMBER OF SEQ ID NOS: 415
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 341
;   LENGTH: 702
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-157-031-341

  Query Match             5.9%;  Score 137.5;  DB 14;  Length 702;
  Best Local Similarity   18.8%;  Pred. No. 0.081;
  Matches   96;  Conservative   76;  Mismatches  215;  Indels  123;  Gaps   20;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
Db        147 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV------------------ 187

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
Db        188 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 245

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
Db        246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 303

Qy        189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
Db        304 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 363

Qy        247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
Db        364 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 423

Qy        288 LPQYAGSGNLTLALEAKT----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
Db        424 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 483

Qy        335 SPKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLE 381
Db        484 SGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ---- 539

Qy        382 SNIKVLPTWSTPVHPRASALPAPPT----------------------GSALPDPQTASAL 419
Db        540 ---------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVL 590

Qy        420 --PDPPAASALPAALAVISFLLGLGLGVAC 447
Db        591 YGPDTPIISPPDS-----SYLSGANLNLSC 615


RESULT 93
US-10-207-655-87
; Sequence 87, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
;  APPLICANT: Ledbetter, Jeffrey A.
;  APPLICANT:  Hayden-Ledbetter, Martha S.
;  TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
;  FILE REFERENCE: 390069.401C1
;  CURRENT APPLICATION NUMBER: US/10/207,655
;  CURRENT FILING DATE:  2002-07-25
;  NUMBER OF SEQ ID NOS: 426
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 87
;   LENGTH: 702
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-207-655-87

  Query Match             5.9%;  Score 137.5;  DB 14;  Length 702;
  Best Local Similarity   18.8%;  Pred. No. 0.081;
  Matches   96;  Conservative   76;  Mismatches  215;  Indels  123;  Gaps   20;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
Db        147 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV------------------ 187

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
Db        188 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 245

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
Db        246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 303

Qy        189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
Db        304 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 363

Qy        247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
Db        364 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 423

Qy        288 LPQYAGSGNLTLALEAKT----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
Db        424 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 483

Qy        335 SPKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLE 381
Db        484 SGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ---- 539

Qy        382 SNIKVLPTWSTPVHPRASALPAPPT----------------------GSALPDPQTASAL 419
Db        540 ---------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVL 590

Qy        420 --PDPPAASALPAALAVISFLLGLGLGVAC 447
Db        591 YGPDTPIISPPDS-----SYLSGANLNLSC 615


RESULT 94
US-10-117-937-592
; Sequence 592, Application US/10117937
; Publication No. US20030220239A1
; GENERAL INFORMATION:
;  APPLICANT: CTL IMMUNO THERAPIES CORP.
;  APPLICANT:  SIMARD, John, J.L.
;  APPLICANT:  DIAMOND, David, C.
;  APPLICANT:  LIU, Liping
```

```
;  APPLICANT:  XIE, Zhidong
;  TITLE OF INVENTION: EPITOPE SEQUENCES
;  FILE REFERENCE: CTLIMM.027A
;  CURRENT APPLICATION NUMBER: US/10/117,937
;  CURRENT FILING DATE:  2002-04-04
;  PRIOR APPLICATION NUMBER: US 60/282,211
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: US 60/337,017
;  PRIOR FILING DATE: 2001-11-07
;  PRIOR APPLICATION NUMBER: US 60/363,210
;  PRIOR FILING DATE: 2002-03-07
;  NUMBER OF SEQ ID NOS: 602
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 592
;   LENGTH: 702
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-117-937-592

  Query Match             5.9%;  Score 137.5;  DB 15;  Length 702;
  Best Local Similarity   18.8%;  Pred. No. 0.081;
  Matches   96;  Conservative   76;  Mismatches  215;  Indels  123;  Gaps   20;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
                |: :  |   :  | | |  ||    : :    | |:  : :
Db        147 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV------------------ 187

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
               | |     ||  | : |:   |: :| || ::     : : | | :::|  | :
Db        188 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 245

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
               :|    |::| |:  :   |:|  |       |   | : | :  : : :||::||
Db        246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 303

Qy        189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
               :  :    :  : :  :      :        ||    :| | | ::  :  |||   ::
Db        304 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 363

Qy        247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
                    :     |: :::  ||::          :| :    |: |           |: :
Db        364 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 423

Qy        288 LPQYAGSGNLTLALEAKT-----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
                 |    ||:|:  | :            | : |    | :   |:    | ||:
Db        424 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 483

Qy        335 SPKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLE 381
              |      ::|      |::      |   |||       :   |||      |   :||
Db        484 SGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ---- 539

Qy        382 SNIKVLPTWSTPVHPRASALPAPPT---------------------GSALPDPQTASAL 419
                  | || ||          |                       :   || |   |
Db        540 ---------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVL 590

Qy        420 --PDPPAASALPAALAVISFLLGLGLGVAC 447
                ||  |  |   :     |:| |  | ::|
Db        591 YGPDTPIISPPDS-----SYLSGANLNLSC 615


RESULT 95
US-10-295-027-332
; Sequence 332, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
;  APPLICANT: Afar, Daniel
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Ginsberg, Wendy M.
;  APPLICANT:  Gish, Kurt C.
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Hevezi, Peter A.
;  APPLICANT:  Mack, David H.
;  APPLICANT:  Murray, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Modulators of Cancer
;  FILE REFERENCE: 018501-012500US
;  CURRENT APPLICATION NUMBER: US/10/295,027
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: US 09/663,733
;  PRIOR FILING DATE: 2000-09-15
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/335,394
;  PRIOR FILING DATE: 2001-11-15
;  PRIOR APPLICATION NUMBER: US 60/332,464
;  PRIOR FILING DATE: 2001-11-21
;  PRIOR APPLICATION NUMBER: US 60/334,393
;  PRIOR FILING DATE: 2001-11-29
;  PRIOR APPLICATION NUMBER: US 60/340,376
;  PRIOR FILING DATE: 2001-12-14
;  PRIOR APPLICATION NUMBER: US 60/347,211
;  PRIOR FILING DATE: 2002-01-08
;  PRIOR APPLICATION NUMBER: US 60/347,349
;  PRIOR FILING DATE: 2002-01-10
;  PRIOR APPLICATION NUMBER: US 60/355,250
;  PRIOR FILING DATE: 2002-02-08
;  PRIOR APPLICATION NUMBER: US 60/356,714
;  PRIOR FILING DATE: 2002-02-13
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 1386
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 332
;   LENGTH: 702
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-295-027-332

  Query Match             5.9%;  Score 137.5;  DB 15;  Length 702;
  Best Local Similarity   18.8%;  Pred. No. 0.081;
  Matches   96;  Conservative   76;  Mismatches  215;  Indels  123;  Gaps   20;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
                |: :  |   :  | | |  ||    : :    | |:  : :
Db        147 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV------------------ 187

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
               | |     ||  | : |:   |: :| || ::     : : | | :::|  | :
Db        188 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 245

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
               :|    |::| |:  :   |:|  |       |   | : | :  : : :||::||
Db        246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 303

Qy        189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
               :  :    :  : :  :      :        ||    :| | | ::  :  |||   ::
Db        304 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 363

Qy        247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
                    :     |: :::  ||::          :| :    |: |           |: :
Db        364 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 423

Qy        288 LPQYAGSGNLTLALEAKT-----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
                 |    ||:|:  | :            | : |    | :   |:    | ||:
Db        424 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 483

Qy        335 SPKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLE 381
              |      ::|      |::      |   |||       :   |||      |   :||
Db        484 SGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ---- 539

Qy        382 SNIKVLPTWSTPVHPRASALPAPPT---------------------GSALPDPQTASAL 419
                  | || ||          |                       :   || |   |
Db        540 ---------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVL 590

Qy        420 --PDPPAASALPAALAVISFLLGLGLGVAC 447
```

```
                 || |  |   :     |:| |  | ::|
Db        591 YGPDTPIISPPDS-----SYLSGANLNLSC 615


RESULT 96
US-10-245-871-282
; Sequence 282, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
;  APPLICANT: HUMPHREYS, ROBERT
;  APPLICANT:  XU, MINZHEN
;  TITLE OF INVENTION: Ii-KEY/ANTIGENIC EPITOPE HYBRID PEPTIDE VACCINES
;  FILE REFERENCE: REH-2013
;  CURRENT APPLICATION NUMBER: US/10/245,871
;  CURRENT FILING DATE:  2003-01-09
;  PRIOR APPLICATION NUMBER: 10/197,000
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: 09/396,813
;  PRIOR FILING DATE: 1999-09-14
;  NUMBER OF SEQ ID NOS: 905
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 282
;   LENGTH: 702
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-245-871-282

  Query Match             5.9%;  Score 137.5;  DB 15;  Length 702;
  Best Local Similarity   18.8%;  Pred. No. 0.081;
  Matches   96;  Conservative   76;  Mismatches  215;  Indels  123;  Gaps   20;

Qy          20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
               |: : |   :  | | |  ||    : :    | |: : :
Db         147 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV------------------ 187

Qy          80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
                 | |      ||  | : |:   |: :| || ::     : : | | :::|  | :
Db         188 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 245

Qy         130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
               :|     |::| |:  :   |:|  |       |   |  : | :  : : :||::||
Db         246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 303

Qy         189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
               :  :  :  : :  :    :        ||    :| | | ::  :  |||   ::
Db         304 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 363

Qy         247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
                     :     |: :::  ||::           :| :    |: |         |:  :
Db         364 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 423

Qy         288 LPQYAGSGNLTLALEAKT-----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
                 |    ||:|:  | :           | : |    | :   |:    | ||:
Db         424 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 483

Qy         335 SPKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLE 381
               |     ::|     |::      |   |||       :   |||     |   :||
Db         484 SGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ---- 539

Qy         382 SNIKVLPTWSTPVHPRASALPAPPT---------------------GSALPDPQTASAL 419
                   | || ||         |                        :   || |   |
Db         540 ---------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVL 590

Qy         420 --PDPPAASALPAALAVISFLLGLGLGVAC 447
                 || |  |   :     |:| |  | ::|
Db         591 YGPDTPIISPPDS-----SYLSGANLNLSC 615


RESULT 97
US-09-756-551A-17
; Sequence 17, Application US/09756551A
; Patent No. US20020051768A1
;  GENERAL INFORMATION:
;    APPLICANT:  C. Morrow et al.
;    TITLE OF INVENTION:  ENCAPSIDATED RECOMBINANT VIRAL
;    TITLE OF INVENTION:  NUCLEIC ACID AND METHODS OF MAKING AND
;    TITLE OF INVENTION:  USING SAME
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 STATE STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/756,551A
;      FILING DATE:  08-JAN-2001
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/376,184
;      FILING DATE:  17-AUG-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/987,867
;      FILING DATE:  09-DEC-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/389,459
;      FILING DATE:  15-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/087,009
;      FILING DATE:  01-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lauro, Peter C.
;      REGISTRATION NUMBER:  32,360
;      REFERENCE/DOCKET NUMBER:  UAI-004CPDV2CN
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 742-4214
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  734 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-756-551A-17

  Query Match             5.9%;  Score 137.5;  DB 9;  Length 734;
  Best Local Similarity   18.8%;  Pred. No. 0.086;
  Matches   96;  Conservative   76;  Mismatches  215;  Indels  123;  Gaps   20;

Qy          20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
               |: : |   :  | | |  ||    : :    | |: : :
Db         179 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV------------------ 219

Qy          80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
                 | |      ||  | : |:   |: :| || ::     : : | | :::|  | :
Db         220 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 277

Qy         130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
               :|     |::| |:  :   |:|  |       |   |  : | :  : : :||::||
Db         278 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 335

Qy         189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
               :  :  :  : :  :    :        ||    :| | | ::  :  |||   ::
Db         336 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 395

Qy         247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
                     :     |: :::  ||::           :| :    |: |         |:  :
Db         396 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 455
```

```
Qy        288 LPQYAGSGNLTLALEAKT-----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
Db        456 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 515

Qy        335 SPKLMLSLKLENKEAKV------SKREKPV-------WVLNPEAGMWQCLLSDSGQVLLE 381
Db        516 SGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ---- 571

Qy        382 SNIKVLPTWSTPVHPRASALPAPPT----------------------GSALPDPQTASAL 419
Db        572 ---------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVL 622

Qy        420 --PDPPAASALPAALAVISFLLGLGLGVAC 447
Db        623 YGPDTPIISPPDS-----SYLSGANLNLSC 647


RESULT 98
US-10-451-168-78
; Sequence 78, Application US/10451168
; Publication No. US20040091969A1
; GENERAL INFORMATION:
;  APPLICANT: SMITHKLINE BEECHAM CORPORATION
;  APPLICANT:  SMITHKLINE BEECHAM p.l.c.
;  APPLICANT:  GLAXO GROUP LIMITED
;  TITLE OF INVENTION: NOVEL COMPOUNDS
;  FILE REFERENCE: GP50039
;  CURRENT APPLICATION NUMBER: US/10/451,168
;  CURRENT FILING DATE:  2003-11-12
;  PRIOR APPLICATION NUMBER: PCT/US01/49232
;  PRIOR FILING DATE: 2000-12-17
;  PRIOR APPLICATION NUMBER: 60/256,710
;  PRIOR FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: 60/257,048
;  PRIOR FILING DATE: 2000-12-20
;  PRIOR APPLICATION NUMBER: 60/260,482
;  PRIOR FILING DATE: 2001-01-09
;  PRIOR APPLICATION NUMBER: 60/264,922
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: 60/266,797
;  PRIOR FILING DATE: 2001-02-06
;  PRIOR APPLICATION NUMBER: 60/276,988
;  PRIOR FILING DATE: 2001-03-19
;  PRIOR APPLICATION NUMBER: 60/281,535
;  PRIOR FILING DATE: 2001-04-04
;  PRIOR APPLICATION NUMBER: 60/289,622
;  PRIOR FILING DATE: 2002-06-28
;  NUMBER OF SEQ ID NOS: 110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 78
;   LENGTH: 5635
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-451-168-78

  Query Match             5.9%;  Score 137;  DB 16;  Length 5635;
  Best Local Similarity   20.4%;  Pred. No. 1.5;
  Matches   93;  Conservative   54;  Mismatches  170;  Indels  138;  Gaps   21;

Qy         34 GDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRSLWDQGNFP 93
Db        807 GSNVTLPCYVQGYPEPTIKWRRLDNMPIFSR--PFSVSSISQLRTGA------------ 851

Qy         94 LIIKNLKIEDSDTYICEVEDQ----KEEVQLLVFGLTA------NSDTHLLQGQSLTLTL 143
Db        852 LFILNLWASDKGTYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVANVIEGQQLTLPC 911

Qy        144 ESPPGS---------------SPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTC--- 184
Db        912 TLLAGNPIPERRWIKNSAMLLQNPYITVRS--------DGSLHIERVQLQDGGEYTCVAS 963

Qy        185 TVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAER 244
```

```
Db        964 NVAGTNNKTTSVVVHVLPTIQHGQQILSTIEG--IPVTLP----------------CKA 1004

Qy        245 ASSSKSWITFDLKNKEVSVKRVTQDPKLQMGKKLPLHLTLPQALPQYAGSGNLTLALEAK 304
Db       1005 SGNPKPSVIWSKKGELIS----TSSAKFSAGADGSLYVVSPGG----EESGEYVCTATNT 1056

Qy        305 TGKLHQEVNLVV---------MRATQLQK------------NLTCEV---------W--- 331
Db       1057 AGYAKRKVQLTVYVRPRVFGDQRGLSQDKPVEISVLAGEEVTLPCEVKSLPPPIITWAKE 1116

Qy        332 ----GPTSPK--LMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSD-SGQV--LLES 382
Db       1117 TQLISPFSPRHTFLPSGSMKITETRTS----------DSGMYLCVATNIAGNVTQAVKL 1165

Qy        383 NIKVLPTWST-PVHPRASA-----LPAPPTGSALP 411
Db       1166 NVHVPPKIQRGPKHLKVQVGQRVDIPCNAQGTPLP 1200


RESULT 99
US-10-106-698-4871
; Sequence 4871, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
;  FILE REFERENCE: PA005P1
;  CURRENT APPLICATION NUMBER: US/10/106,698
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: PCT/US00/26524
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US 60/157,137
;  PRIOR FILING DATE: 1999-09-29
;  PRIOR APPLICATION NUMBER: US 60/163,280
;  PRIOR FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 8564
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4871
;   LENGTH: 595
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (392)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (393)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (571)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-4871

  Query Match             5.9%;  Score 136.5;  DB 14;  Length 595;
  Best Local Similarity   18.5%;  Pred. No. 0.077;
  Matches   95;  Conservative   75;  Mismatches  214;  Indels  129;  Gaps   20;

Qy         20 PAATQGNKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDR 79
Db         40 PSISSNNSKPVEDK-DAVAFTCEPETQDATYLWWVNNQSLPV------------------ 80

Qy         80 ADSRRSLWDQGNFPLIIKNLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS---- 129
Db         81 --SPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPL 138

Qy        130 DTHLLQGQSLTLTLESPPGSSPSVQCR-SPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQ 188
Db        139 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHN 196

Qy        189 NQKKV-EFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVE-KLTGSGELWWQAERAS 246
Db        197 SDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSL 256
```

```
Qy        247 SSKSWITFDLKNKEVSVKRVTQDP----------KLQMGKKLPLHL---------TLPQA 287
Db        257 PVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPS 316

Qy        288 LPQYAGSGNLTLALEAKT----------GKLHQEV-NLVVMRATQLQKNL-TCEVWGPT 334
Db        317 YTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQANNSA 376

Qy        335 S---------------PKLMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLLSDSGQV 378
Db        377 SGHSRTTVKTITVSAXXPKPSIS---SNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ- 432

Qy        379 LLESNIKVLPTWSTPVHPRASALPAPPT---------------------GSALPDPQTA 416
Db        433 -----------SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTL 480

Qy        417 SAL--PDPPAASALPAALAVISFLLGLGLGVAC 447
Db        481 DVLYGPDTPIISPPDS-----SYLSGANLNLSC 508
```

```
RESULT 100
US-09-808-602-67
; Sequence 67, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
;  APPLICANT: Vernet, Corine A
;  APPLICANT:  Fernandes, Elma
;  APPLICANT:  Shimkets, Richard A
;  APPLICANT:  Herrman, John L
;  APPLICANT:  Majumder, Kumud
;  APPLICANT:  Mishra, Vishnu
;  APPLICANT:  Mezes, Peter S
;  APPLICANT:  MacDougall, John
;  TITLE OF INVENTION: No. US20020155115A1el Proteins and Nuclec Acids Encoding Same
;  FILE REFERENCE: 15966-697 CIP
;  CURRENT APPLICATION NUMBER: US/09/808,602
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: 09/800,198
;  PRIOR FILING DATE: 2001-03-05
;  PRIOR APPLICATION NUMBER: 60/186,596
;  PRIOR FILING DATE: 2000-03-03
;  NUMBER OF SEQ ID NOS: 114
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 67
;   LENGTH: 1395
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-808-602-67

  Query Match             5.9%;  Score 136.5;  DB 9;  Length 1395;
  Best Local Similarity   21.1%;  Pred. No. 0.25;
  Matches  102;  Conservative   53;  Mismatches  179;  Indels  149;  Gaps   18;

Qy         26 NKVVLGKKGDTVELTCTASQKKSIQFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRRS 85
Db        353 NKKV-GLNG-VVQLPCMASGNPPPSVFWTK--------------EGVSTLMFPNSSHGR 395

Qy         86 LWDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLLVFGLTANSDTH---LLQGQSLTLT 142
Db        396 QYVAADGTLQITDVRQEDEGYYVCSAFSVVDSSTVRVFLQVSSVDERPPPIIQIGPANQT 455

Qy        143 LESPPGSSPSVQCR-----SPRGK-------------NIQGGKTLSVSQLELQDSGTWT 183
Db        456 L--PKGSVATLPCRATGNPSPRIKWFHDGHAVQAGNRYSIIQGSSLRVDDLQLSDSGTYT 513

Qy        184 CTVLQNQKKVEFKIDIVVLAFQKASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELWWQAE 243
Db        514 CTA-------------------------SGERGETSWAATLTVEK-PGSTSLHRAAD 544

Qy        244 ----------------RASSSKSWITFDLKNKEVS-VKRVTQD---PKLQMGKKLPLH- 281
```

```
Db        545 PSTYPAPPGTPKVLNVSRTSISLRWAKSQEKPGAVGPIIGYTVEYFSPDLQTGWIVAAHR 604

Qy        282 ---------------------------LTLPQALPQY----------AGSGNLTLALEA 303
Db        605 VGDTQVTISGLTPGTSYVFLVRAENTQGISVPSGLSNVIKTIEADFDAASANDLSAARTL 664

Qy        304 KTGKLHQEVNLVVMRATQLQKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPVWVLNP 363
Db        665 LTGKSVELIDASAINASAVRLEWMLHV--SADEKYVEGLRIHYKDASV-----------P 711

Qy        364 EAGMWQCLLSD-SGQVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALPDP 422
Db        712 SAQYHSITVMDASAESFVVGNLKKYTKYEFFLTPFFETIEGQPSNS-----KTALTYEDV 766

Qy        423 PAA 425
Db        767 PSA 769
```

```
Search completed: August  3, 2004, 13:47:44
Job time : 45.3425 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 3, 2004, 13:00:24 ; Search time 10.4339 Seconds
(without alignments)
4185.504 Million cell updates/sec

Title: SEQ4
Perfect score: 2414
Sequence: 1 MNRGVPFRHLLLVLQLALLP.....DETCAEAQDGELDGLWTTDP 454

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1227.5	50.8	330	1	GHHU	Ig gamma-1 chain C
2	1221	50.6	374	2	S69339	Ig heavy chain V r
3	1203.5	49.9	255	4	S31866	Ig gamma-1 chain C
4	1173.5	48.6	234	2	PT0207	Ig gamma chain C r
5	1157	47.9	377	2	A23511	Ig gamma-3 chain C
6	1155	47.8	377	2	A60764	Ig gamma-3 chain C
7	1133	46.9	289	1	G3HUWI	Ig gamma-3 heavy c
8	1130.5	46.8	326	1	G2HU	Ig gamma-2 chain C
9	1119	46.4	327	1	G4HU	Ig gamma-4 chain C
10	1023	42.4	458	1	RWHUT4	T-cell surface gly
11	955.5	39.6	398	1	G3MSM	Ig gamma-3 chain C
12	932.5	38.6	393	1	G1MSM	Ig gamma-1 chain C
13	915.5	37.9	470	2	S22080	Ig heavy chain pre
14	906.5	37.6	399	1	G2MSAM	Ig gamma-2a chain
15	901	37.3	328	2	I47159	Ig gamma 2a chain
16	899	37.2	323	1	GHRB	Ig gamma chain C r
17	898	37.2	277	2	I47162	Ig gamma 4 chain c
18	898	37.2	328	2	I47160	Ig gamma 2b chain
19	894.5	37.1	329	1	G2GP	Ig gamma-2 chain C
20	879.5	36.4	328	2	I47158	Ig gamma 1 chain c
21	877.5	36.4	405	1	G2MSBM	Ig gamma-2b chain
22	875	36.2	432	1	RWCZT4	T-cell surface gly
23	873	36.2	328	2	I47161	Ig gamma 3 chain c
24	857.5	35.5	329	1	G3MSC	Ig gamma-3 chain C
25	854.5	35.4	444	2	PC4436	monoclonal antibod
26	854.5	35.4	472	2	S31459	Ig gamma-1 chain -
27	850.5	35.2	469	2	S37483	Ig gamma-2a chain
28	838	34.7	308	2	C30554	Ig heavy chain C r
29	833	34.5	333	2	PS0018	Ig gamma-2b chain
30	825	34.2	446	2	S40295	Ig gamma-2a chain
31	823.5	34.1	326	2	PS0017	Ig gamma-1 chain C
32	823	34.1	329	2	S00847	Ig gamma-2c chain
33	818.5	33.9	324	1	G1MS	Ig gamma-1 chain C
34	818	33.9	474	1	G2MS11	Ig gamma-2b chain
35	805.5	33.4	330	1	G2MSA	Ig gamma-2a chain
36	801	33.2	335	1	G2MSAB	Ig gamma-2a chain
37	793.5	32.9	475	2	S01321	Ig gamma-2b chain
38	792.5	32.8	322	2	PS0019	Ig gamma-2a chain
39	790	32.7	432	1	RWMQT4	T-cell surface gly
40	774	32.1	327	2	S06611	Ig gamma-2 chain C
41	702	29.1	180	2	I46732	Ig gamma heavy cha
42	601.5	24.9	459	2	A46254	CD4 precursor - ra
43	574.5	23.8	218	2	A36040	Ig heavy chain V-I
44	574.5	23.8	432	2	S30193	T-cell surface gly
45	572.5	23.7	249	2	S69340	Ig heavy chain VHI
46	566	23.4	152	2	S14236	Ig gamma-1 chain C
47	495	20.5	457	2	A27449	T-cell surface gly
48	475	19.7	457	1	RWMST4	T-cell surface gly
49	411.5	17.0	572	2	B46529	Ig Y heavy chain (
50	402.5	16.7	549	2	S04845	Ig heavy chain pre
51	389	16.1	627	2	S14683	Ig mu chain precur
52	388	16.1	388	1	EHMS	Ig epsilon chain C
53	386.5	16.0	474	2	S15590	Ig heavy chain - h
54	385	15.9	453	2	S37768	Ig mu chain C regi
55	384	15.9	592	2	S25705	Ig mu chain - shee
56	381	15.8	473	1	MHHUM	Ig mu chain C regi
57	381	15.8	548	2	S38864	Ig epsilon chain C
58	376.5	15.6	452	1	MHHU	Ig mu chain C regi
59	374	15.5	448	2	S03186	Ig heavy chain C r
60	373.5	15.5	476	1	MHMSM	Ig mu chain C regi
61	370.5	15.3	429	1	EHRT	Ig epsilon chain C
62	369.5	15.3	479	1	MHRBM	Ig mu chain C regi
63	369	15.3	450	1	MHDG	Ig mu chain C regi
64	365.5	15.1	391	1	MHHUBT	Ig mu heavy chain
65	363.5	15.1	423	1	EHMSS	Ig epsilon chain C
66	363	15.0	455	2	A24976	Ig mu chain C regi
67	362	15.0	455	1	MHMS	Ig mu chain C regi
68	361.5	15.0	426	2	I36948	Ig epsilon-chain -
69	361.5	15.0	428	1	EHHU	Ig epsilon chain C
70	361	15.0	504	2	S00390	Ig gamma chain (cl
71	360	14.9	458	1	MHRB	Ig mu chain C regi
72	357	14.8	71	2	I60082	CD4 receptor - hum
73	354.5	14.7	457	2	S03961	Ig mu chain C regi
74	353	14.6	343	2	S25644	Ig mu chain C regi
75	349	14.5	454	1	MHHY	Ig mu chain C regi
76	345.5	14.3	453	2	C31933	Ig mu chain C regi
77	331.5	13.7	112	2	B30503	Ig gamma-2a chain
78	325.5	13.5	577	2	I50731	Ig heavy chain - n
79	321.5	13.3	433	2	S31436	Ig upsilon chain -
80	312	12.9	438	1	HVRKC2	Ig mu chain C regi
81	308.5	12.8	474	2	I50830	Ig mu chain - Lepi
82	305.5	12.7	99	2	S21461	T-cell surface gly
83	301.5	12.5	585	2	A46507	Ig alpha chain - c
84	301.5	12.5	684	2	S60266	novel antigen rece
85	300	12.4	342	2	I47175	Ig alpha chain C r
86	297	12.3	339	2	S09264	Ig alpha chain C r
87	296	12.3	357	2	S09269	Ig alpha chain C r
88	296	12.3	367	1	MHCH	Ig mu chain C regi
89	296	12.3	454	2	A46532	Ig mu chain C regi
90	295.5	12.2	343	2	S09272	Ig alpha chain C r
91	295.5	12.2	461	1	HVRKC0	Ig mu chain C regi
92	294.5	12.2	340	2	B22360	Ig alpha-2 chain C
93	294	12.2	438	1	HVRKCS	Ig mu chain C regi
94	293	12.1	299	1	AHRB	Ig alpha chain C r
95	292	12.1	338	2	S09276	Ig alpha chain C r
96	292	12.1	352	2	S09266	Ig alpha chain C r
97	290	12.0	568	2	A34891	Ig heavy chain pre
98	289	12.0	357	2	S09265	Ig alpha chain C r
99	286.5	11.9	340	2	I56230	Ig alpha-2 chain -
100	286.5	11.9	370	1	HVRKC5	Ig mu chain C regi
101	285.5	11.8	348	2	S09270	Ig alpha chain C r
102	284	11.8	244	2	S12328	Ig heavy chain C r

103	282.5	11.7	348	2	S09273	Ig alpha chain C r
104	282	11.7	347	2	S09274	Ig alpha chain C r
105	281	11.6	393	1	HVRKC1	Ig mu chain C regi
106	280.5	11.6	99	2	S21462	T-cell surface gly
107	279	11.6	352	2	S05500	Ig alpha-1 chain C
108	277.5	11.5	353	1	A1HU	Ig alpha-1 chain C
109	277	11.5	340	1	A2HU	Ig alpha-2 chain C
110	275.5	11.4	88	2	A30503	Ig gamma-2b chain
111	275	11.4	357	2	S09267	Ig alpha chain C r
112	275	11.4	568	2	A45804	Ig mu chain C regi
113	268.5	11.1	360	2	S09271	Ig alpha chain C r
114	264.5	11.0	344	1	AHMS	Ig alpha chain C r
115	263.5	10.9	358	2	S09268	Ig alpha chain C r
116	258	10.7	580	2	A46538	Ig heavy chain, se
117	257.5	10.7	342	2	A45966	Ig alpha chain C r
118	257	10.6	111	2	S43148	Ig upsilon chain -
119	253.5	10.5	573	2	S12838	Ig mu chain precur
120	245	10.1	335	2	S09275	Ig alpha chain C r
121	243.5	10.1	220	2	C22360	Ig alpha-2 chain C
122	241.5	10.0	448	2	A46533	Ig heavy chain C r
123	223.5	9.3	402	2	S20002	Ig heavy chain, se
124	223.5	9.3	1005	2	T18537	Ig heavy chain - c
125	222.5	9.2	228	2	S03050	Ig gamma chain (cl

ALIGNMENTS

RESULT 1
GHHU
Ig gamma-1 chain C region - human
C;Species: Homo sapiens (man)
C;Date: 31-Jan-1981 #sequence_revision 18-Aug-1982 #text_change 16-Jul-1999
C;Accession: A93433; S36861; S33887; B90563; A90564; B91668; A91723; A02146
R;Ellison, J.W.; Berson, B.J.; Hood, L.E.
Nucleic Acids Res. 10, 4071-4079, 1982
A;Title: The nucleotide sequence of a human immunoglobulin C-gamma1 gene.
A;Reference number: A93433; MUID:82274238; PMID:6287432
A;Accession: A93433
A;Molecule type: DNA
A;Residues: 1-330 <ELL>
A;Cross-references: EMBL:Z17370
A;Note: this sequence has the G1m(17) allotypic marker, 97-Lys, and the G1m(1) markers,
A;Note: Lys-330 is removed after translation
R;Harris, L.J.
submitted to the EMBL Data Library, October 1992
A;Reference number: S33904
A;Accession: S36861
A;Molecule type: DNA
A;Residues: 2-330 <HAR>
A;Cross-references: EMBL:Z17370
R;Takahashi, N.; Ueda, S.; Obata, M.; Nikaido, T.; Nakai, S.; Honjo, T.
Cell 29, 671-679, 1982
A;Title: Structure of human immunoglobulin gamma genes: implications for evolution of a
A;Reference number: S33887; MUID:83001943; PMID:6811139
A;Accession: S33887
A;Molecule type: DNA
A;Residues: 88-113;235-330 <TAK>
A;Cross-references: EMBL:Z17370
R;Cunningham, B.A.; Rutishauser, U.; Gall, W.E.; Gottlieb, P.D.; Waxdal, M.J.; Edelman,
Biochemistry 9, 3161-3170, 1970
A;Title: The covalent structure of a human gammaG-immunoglobulin. VII. Amino acid sequen
A;Reference number: A90563; MUID:71064024; PMID:5489771
A;Contents: myeloma protein Eu
A;Accession: B90563
A;Molecule type: protein
A;Residues: 1-96,'R',98-135 <CUN>
A;Note: this sequence has the G1m(3) marker, 97-Arg
R;Rutishauser, U.; Cunningham, B.A.; Bennett, C.; Konigsberg, W.H.; Edelman, G.M.
Biochemistry 9, 3171-3181, 1970
A;Title: The covalent structure of a human gammaG-immunoglobulin. VIII. Amino acid seque
A;Reference number: A90564; MUID:71064025; PMID:5530842
A;Contents: Eu
A;Accession: A90564
A;Molecule type: protein
A;Residues: 136-154,'Q',156-165,'Q',167-176,'Q',178-194,'N',196-197,'D',199-238,'E',240,'
A;Note: this sequence has the G1m(non-1) markers, 239-Glu and 241-Met
R;Ponstingl, H.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 357, 1571-1604, 1976
A;Title: Die Primaerstruktur eines monoklonalen IgG1-Immunglobulins (Myelomprotein Nie),
igen Primaerstruktur.
A;Reference number: A91668; MUID:77070269; PMID:826475
A;Contents: myeloma protein Nie
A;Accession: B91668
A;Molecule type: protein
A;Residues: 1-34,'Q',36-96,'K',98-115,'Q',117-197,'D',199-238,'D',240,'L',242-268,'E',27(
A;Note: this sequence has the G1m(17) and G1m(1) markers
R;Schmidt, W.E.; Jung, H.D.; Palm, W.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 364, 713-747, 1983
A;Title: Die Primaerstruktur des kristallisierbaren monoklonalen Immunglobulins IgG1 KOL,
A;Reference number: A91723; MUID:83289131; PMID:6884994
A;Contents: myeloma protein KOL; disulfide bonds
A;Accession: A91723
A;Molecule type: protein
A;Residues: 1-96,'R',98-197,'D',199-238,'E',240,'M',242-266,'D',268-271,'D',273-330 <SCH:
A;Note: this sequence has the G1m(3) and G1m(non-1) markers
R;Gall, W.E.; Edelman, G.M.
Biochemistry 9, 3188-3196, 1970
A;Title: The covalent structure of a human gammaG-immunoglobulin. X. Intrachain disulfide
A;Reference number: A90565; MUID:71064027; PMID:4923144
A;Contents: annotation; disulfide bonds
R;Dreker, L.; Schwarz, J.; Reichel, W.; Hilschmann, N.
Hoppe-Seyler's Z. Physiol. Chem. 357, 1515-1540, 1976
A;Title: Rule of antibody structure. The primary structure of monoclonal IgG1 immunoglobu
enbromide cleavage products, and the disulfide bridges.
A;Reference number: A91667; MUID:77070267; PMID:1002129
A;Contents: annotation; disulfide bonds
C;Genetics:
A;Gene: GDB:IGHG1
A;Cross-references: GDB:120085; OMIM:147100
A;Map position: 14q32.33-14q32.33
A;Introns: 99/1; 114/1; 224/1
C;Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kapp
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la
C;Superfamily: immunoglobulin C region; immunoglobulin homology
C;Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin
F;20-85/Domain: immunoglobulin homology <IM1>
F;137-206/Domain: immunoglobulin homology <IM2>
F;243-310/Domain: immunoglobulin homology <IM3>
F;27-83,144-204,250-308/Disulfide bonds: #status experimental
F;103/Disulfide bonds: interchain (to light chain) #status experimental
F;109,112/Disulfide bonds: interchain (to heavy chain) #status experimental
F;180/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 50.8%; Score 1227.5; DB 1; Length 330;
Best Local Similarity 74.5%; Pred. No. 9.3e-68;
Matches 246; Conservative 7; Mismatches 22; Indels 55; Gaps 6;

```
Qy      151 PSVQCRSPRGKNIQGG-------------KTLSVS-------------QLELQDSG---- 180
            |||   :|  |:  ||              : ::||                || ||
Db        6 PSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSL 65

Qy      181 --------------TWTCTVLQNQKKVEFKIDIVPCPAPEPKSCDKTHTC-----PELLG 221
                         |: | |  | |    |:|       ||||||||||      |||||
Db       66 SSVVTVPSSSLGTQTYICNV--NHKPSNTKVD----KKVEPKSCDKTHTCPPCPAPELLG 119

Qy      222 GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQY 281
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      120 GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQY 179

Qy      282 NSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 341
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      180 NSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD 239

Qy      342 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSR 401
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        240 ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSR 299

Qy        402 WQQGNVFSCSVMHEALHNHYTQKSLSLSPG 431
              ||||||||||||||||||||||||||||||
Db        300 WQQGNVFSCSVMHEALHNHYTQKSLSLSPG 329


RESULT 2
S69339
Ig heavy chain V region precursor - human
C;Species: Homo sapiens (man)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 01-Dec-2000
C;Accession: S69339; S72664
R;Khamlichi, A.A.; Aucouturier, P.; Preud'homme, J.L.; Cogne, M.
Eur. J. Biochem. 229, 54-60, 1995
A;Title: Structure of abnormal heavy chains in human heavy-chain-deposition disease.
A;Reference number: S69339; MUID:95262687; PMID:7744049
A;Accession: S69339
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-374 <KHA>
A;Cross-references: EMBL:X81695
R;Khamlichi, A.A.
submitted to the EMBL Data Library, September 1994
A;Reference number: S72664
A;Accession: S72664
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-140,'C',142-374 <KH2>
A;Cross-references: EMBL:X81695
C;Superfamily: immunoglobulin C region; immunoglobulin homology

  Query Match             50.6%;  Score 1221;  DB 2;  Length 374;
  Best Local Similarity   71.4%;  Pred. No. 2.7e-67;
  Matches  242;  Conservative   19;  Mismatches   34;  Indels   44;  Gaps    5;

Qy        137 QSLTLT----------------LESPPGS---------------SPSVQCRSPRGKNI 163
              |:||||                :  |||                |||:: |    |:
Db         35 QTLTLTCTFSGFSLSKSGVGVGWIRQPPGQALEWLALIFWDDDKRYSPSLRTRLTITKDT 94

Qy        164 QGGK-TLSVSQLELQDSGTWTCTVL-----QNQKKVEFKIDIVPCPAPEPKSCDKTHTC- 216
              :  |::: ::  |: |: |       |  :   :    :   : ||||||||||
Db         95 SKNQVVLTMTNVDPADTATYYCGYSVEGYGQGYRFHSWGQGTLVTVSSEPKSCDKTHTCP 154

Qy        217 ----PELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNA 272
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 PCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNA 214

Qy        273 KTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQ 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 KTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQ 274

Qy        333 VYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLY 392
              ||||||||:|:||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 VYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLY 334

Qy        393 SKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG 431
              |||||||||||||||||||||||||||||||||||||||
Db        335 SKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG 373


RESULT 3
S31866
Ig gamma-1 chain C region - synthetic
C;Species: synthetic
A;Note: Homo sapiens (man) gene engineered and expressed in Escherichia coli
C;Date: 06-Jan-1995 #sequence_revision 17-Mar-1997 #text_change 19-May-2000
C;Accession: S31866
R;Filpula, D.
submitted to the EMBL Data Library, February 1993
A;Description: Screeing method for protein-protein interactions of cloned gene products.
A;Reference number: S31866
A;Accession: S31866
A;Molecule type: mRNA
A;Residues: 1-255 <FIL>
A;Cross-references: EMBL:X70421; NID:g33068; PIDN:CAA49866.1; PID:g33069
C;Keywords: immunoglobulin
F;1-22/Region: Escherichia coli outer membrane protein A precursor
F;23-255/Region: human Ig gamma-1 chain C region

  Query Match             49.9%;  Score 1203.5;  DB 4;  Length 255;
  Best Local Similarity   97.4%;  Pred. No. 2e-66;
  Matches  225;  Conservative    0;  Mismatches    1;  Indels    5;  Gaps    1;

Qy        206 EPKSCDKTHTC-----PELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 260
              | |||||||||     ||||||||||||||||||||||||||||||||||||||||||||
Db         24 ESKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 83

Qy        261 NWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKT 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 NWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKT 143

Qy        321 ISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTP 380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 ISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTP 203

Qy        381 PVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG 431
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 PVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG 254


RESULT 4
PT0207
Ig gamma chain C region - chimpanzee
C;Species: Pan troglodytes (chimpanzee)
C;Date: 23-Nov-1991 #sequence_revision 23-Nov-1991 #text_change 16-Jul-1999
C;Accession: PT0207
R;Ehrlich, P.H.; Moustafa, Z.A.; Oestberg, L.
Mol. Immunol. 28, 319-322, 1991
A;Title: Nucleotide sequence of chimpanzee Fc and hinge regions.
A;Reference number: PT0207; MUID:91287716; PMID:2062315
A;Accession: PT0207
A;Molecule type: mRNA
A;Residues: 1-234 <EHR>
C;Superfamily: immunoglobulin C region; immunoglobulin homology
C;Keywords: immunoglobulin
F;48-117/Domain: immunoglobulin homology <IMM>

  Query Match             48.6%;  Score 1173.5;  DB 2;  Length 234;
  Best Local Similarity   92.5%;  Pred. No. 1.2e-64;
  Matches  221;  Conservative    3;  Mismatches    2;  Indels   13;  Gaps    2;

Qy        192 KVEFKIDIVPCPAPEPKSCDKTHTC-----PELLGGPSVFLFPPKPKDTLMISRTPEVTC 246
              ||: |:        ||||||  ||||     ||||||||||||||||||||||||||||
Db          4 KVDKKV--------EPKSCDTTHTCPPCAAPELLGGPSVFLFPPKPKDTLMISRTPEVTC 55

Qy        247 VVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKC 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         56 VVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKC 115

Qy        307 KVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEW 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 KVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEW 175

Qy        367 ESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKS 425
              ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 ESSGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKS 234


RESULT 5
A23511
Ig gamma-3 chain C region (allotype G3m(b)) - human
C;Species: Homo sapiens (man)
C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 23-Jul-1999
C;Accession: A23511
R;Huck, S.; Fort, P.; Crawford, D.H.; Lefranc, M.P.; Lefranc, G.
```

```
Nucleic Acids Res. 14, 1779-1789, 1986
A;Title: Sequence of a human immunoglobulin gamma 3 heavy chain constant region gene: cc
A;Reference number: A23511; MUID:86148507; PMID:3081877
A;Accession: A23511
A;Molecule type: DNA
A;Residues: 1-377 <HUC>
A;Cross-references: GB:X03604; GB:M12958; NID:g33070; PIDN:CAA27268.1; PID:g577056
C;Genetics:
A;Gene: GDB:IGHG3
A;Cross-references: GDB:119339; OMIM:147120
A;Map position: 14q32.33-14q32.33
A;Introns: 98/3; 115/3; 130/3; 145/3; 160/3; 270/3
C;Superfamily: immunoglobulin C region; immunoglobulin homology
C;Keywords: immunoglobulin
F;20-85/Domain: immunoglobulin homology <IMM>

  Query Match             47.9%;  Score 1157;  DB 2;  Length 377;
  Best Local Similarity   57.3%;  Pred. No. 2.1e-63;
  Matches  239;  Conservative   26;  Mismatches   54;  Indels   98;  Gaps   10;

Qy         70 TKGPSKLNDRADSRRS--------------------LWDQG-------NFPLIIKNLKI 101
              |||||       || :                    |: |       || ::::  :
Db          3 TKGPSVFPLAPCSRSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGL 62

Qy        102 ED-------------SDTYICEVE--------DQKEEVQLLVFGLTANSDTHLLQGQSLT 140
                           : || | |          |:: |::            ||
Db         63 YSLSSVVTVPSSSLGTQTYTCNVNHKPSNTKVDKRVELK-----TPLGDTTHTCPRCPEP 117

Qy        141 LTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIV 200
               : ::||   |  :|  |:                  :|
Db        118 KSCDTPP---PCPRCPEPK----------------------SCDTPP------------ 139

Qy        201 PCP-APEPKSCDKTHTC-----PELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHE 254
              |||  |||||||    |     |||||||||||||||||||||||||||||||||||||
Db        140 PCPRCPEPKSCDTPPPCPRCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHE 199

Qy        255 DPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALP 314
              ||||:| |||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db        200 DPEVQFKWYVDGVEVHNAKTKPREEQYNSTFRVVSVLTVLHQDWLNGKEYKCKVSNKALP 259

Qy        315 APIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPEN 374
              |||||||| |||||||||||||||:|:|||||||||||||||||||||||:|||||
Db        260 APIEKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESSGQPEN 319

Qy        375 NYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG 431
              || ||||:|||||||||||||||||||||:|||||||||||| :||||||||||
Db        320 NYNTTPPMLDSDGSFFLYSKLTVDKSRWQQGNIFSCSVMHEALHNRFTQKSLSLSPG 376


RESULT 6
A60764
Ig gamma-3 chain C region, form LAT - human
C;Species: Homo sapiens (man)
C;Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 16-Jul-1999
C;Accession: A60764
R;Huck, S.; Lefranc, G.; Lefranc, M.P.
Immunogenetics 30, 250-257, 1989
A;Title: A human immunoglobulin IGHG3 allele (Gmb0, b1, c3, c5, u) with an IGHG4 convert
A;Reference number: A60764; MUID:90007613; PMID:2571587
A;Accession: A60764
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-377 <HUC>
C;Superfamily: immunoglobulin C region; immunoglobulin homology
C;Keywords: immunoglobulin
F;20-85/Domain: immunoglobulin homology <IMM>

  Query Match             47.8%;  Score 1155;  DB 2;  Length 377;
  Best Local Similarity   57.3%;  Pred. No. 2.8e-63;
  Matches  239;  Conservative   26;  Mismatches   54;  Indels   98;  Gaps   10;

Qy         70 TKGPSKLNDRADSRRS--------------------LWDQG-------NFPLIIKNLKI 101
              |||||       || :                    |: |       || ::::  :

Db          3 TKGPSVFPLAPCSRSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGL 62

Qy        102 ED-------------SDTYICEVE--------DQKEEVQLLVFGLTANSDTHLLQGQSLT 140
                           : || | |          |:: |::            ||
Db         63 YSLSSVVTVPSSSLGTQTYTCNVNHKPSNTKVDKRVELK-----TPLGDTTHTCPRCPEP 117

Qy        141 LTLESPPGSSPSVQCRSPRGKNIQGGKTLSVSQLELQDSGTWTCTVLQNQKKVEFKIDIV 200
               : ::||   |  :|  |:                  :|
Db        118 KSCDTPP---PCPRCPEPK----------------------SCDTPP------------ 139

Qy        201 PCP-APEPKSCDKTHTC-----PELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHE 254
              |||  |||||||    |     |||||||||||||||||||||||||||||||||||||
Db        140 PCPRCPEPKSCDTPPPCPRCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHE 199

Qy        255 DPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALP 314
              ||||:| |||||||||||||||||||||:|:|||||||||||||||||||||||:|||||
Db        200 DPEVQFKWYVDGVEVHNAKTKPREEQYNSTFRVVSVLTVLHQDWLNGKEYKCKVSNKALP 259

Qy        315 APIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPEN 374
              |||||||| |||||||||||||||:|:|||||||||||||||||||||||:|||||
Db        260 APIEKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESSGQPEN 319

Qy        375 NYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPG 431
              || ||||:|||||||||||||||||||||:|||||||||||| :||||||||||
Db        320 NYNTTPPMLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNRFTQKSLSLSPG 376


RESULT 7
G3HUWI
Ig gamma-3 heavy chain disease proteins - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1979 #sequence_revision 23-Oct-1981 #text_change 16-Jul-1999
C;Accession: A90442; A92219; A90198; A93915; A02149
R;Frangione, B.; Rosenwasser, E.; Prelli, F.; Franklin, E.C.
Biochemistry 19, 4304-4308, 1980
A;Title: Primary structure of human gamma3 immunoglobulin deletion mutant: gamma3 heavy-c
A;Reference number: A90442; MUID:81021548; PMID:6774747
A;Contents: heavy chain disease protein Wis
A;Accession: A90442
A;Molecule type: protein
A;Residues: 1-289 <FRA>
A;Note: the molecule is a dimer linked by 12 disulfide bonds; it has an extra interchain
A;Note: this protein lacks most of the V region and all of the CH1 region. Residue 12 co
A;Note: the sequence of residues 42-76 was taken from the reference that follows
R;Michaelsen, T.E.; Frangione, B.; Franklin, E.C.
J. Biol. Chem. 252, 883-889, 1977
A;Title: Primary structure of the 'hinge' region of human IgG3. Probable quadruplication
A;Reference number: A92219; MUID:77118561; PMID:402363
A;Contents: normal gamma-3 chains, sequence corresponding to residues 12-97 of protein W
A;Accession: A92219
A;Molecule type: protein
A;Residues: 12-97 <MIC>
A;Note: the hinge region in gamma-3 chains is about four times as long as in other gamma
idue segment (12-28)
A;Note: cysteines at positions 24, 27, 33, 39, 42, 48, 54, 57, 63, 69, and 72 form inter-
R;Wolfenstein-Todel, C.; Frangione, B.; Prelli, F.; Franklin, E.C.
Biochem. Biophys. Res. Commun. 71, 907-914, 1976
A;Title: The amino acid sequence of "heavy chain disease" protein ZUC. Structure of the
A;Reference number: A90198; MUID:77021516; PMID:823945
A;Contents: heavy chain disease protein Zuc, partial sequence corresponding to residues
A;Accession: A90198
A;Molecule type: protein
A;Residues: 59-125,'EB',128-226,228-289 <WOL>
A;Note: this protein lacks most of the V region, all of the CH1 region, and part of the
R;Alexander, A.; Steinmetz, M.; Barritault, D.; Frangione, B.; Franklin, E.C.; Hood, L.;
Proc. Natl. Acad. Sci. U.S.A. 79, 3260-3264, 1982
A;Title: gamma heavy chain disease in man: cDNA sequence supports partial gene deletion
A;Reference number: A93915; MUID:82247835; PMID:6808505
A;Contents: heavy chain disease protein Omm
A;Accession: A93915
A;Molecule type: mRNA
A;Residues: 12-70;72-114;116-125,'E',127-133,'L',135-136,'E',138,'Y',140-154,'D',156-157
A;Note: a carboxyl-terminal Lys is removed posttranslationally
```